Consistency of the mean and the principal components of spatially distributed functional data

Siegfried Hörmann* and Piotr Kokoszka†

Abstract

This paper develops a framework for the estimation of the functional mean and the functional principal components when the functions form a random field. More specifically, the data we study consist of curves $X(\mathbf{s}_k; t)$, $t \in [0, T]$, observed at spatial points $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$. We establish conditions for the sample average (in space) of the $X(\mathbf{s}_k)$ to be a consistent estimator of the population mean function, and for the usual empirical covariance operator to be a consistent estimator of the population covariance operator. These conditions involve an interplay of the assumptions on an appropriately defined dependence between the functions $X(\mathbf{s}_k)$ and the assumptions on the spatial distribution of the points $\mathbf{s}_k$. The rates of convergence may be the same as for iid functional samples, but generally depend on the strength of dependence and appropriately quantified distances between the points $\mathbf{s}_k$. We also formulate conditions for the lack of consistency. The general results are specialized to functional spatial models of practical interest. They are established using an appropriate quadratic loss function which we can bound by terms that reflect the assumptions on the spatial dependence and the distribution of the points. This technique is broadly applicable to all statistics obtained by simple averaging of functional data at spatial locations.

1 Introduction

This paper develops aspects of theory for functional data observed at spatial locations. The data consist of curves $X(\mathbf{s}_k; t)$, $t \in [0, T]$, observed at spatial points $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$. Such data structures are quite common, but often the spatial dependence and the spatial distribution of the points $\mathbf{s}_k$ are not taken into account. A well–known example is the Canadian temperature and precipitation data used as a running example in

---

*Département de Mathémathique, Université Libre de Bruxelles, CP 210, Bd du Triomphe, B-1050 Brussels - Belgium. Email: `shormann@ulb.ac.be`
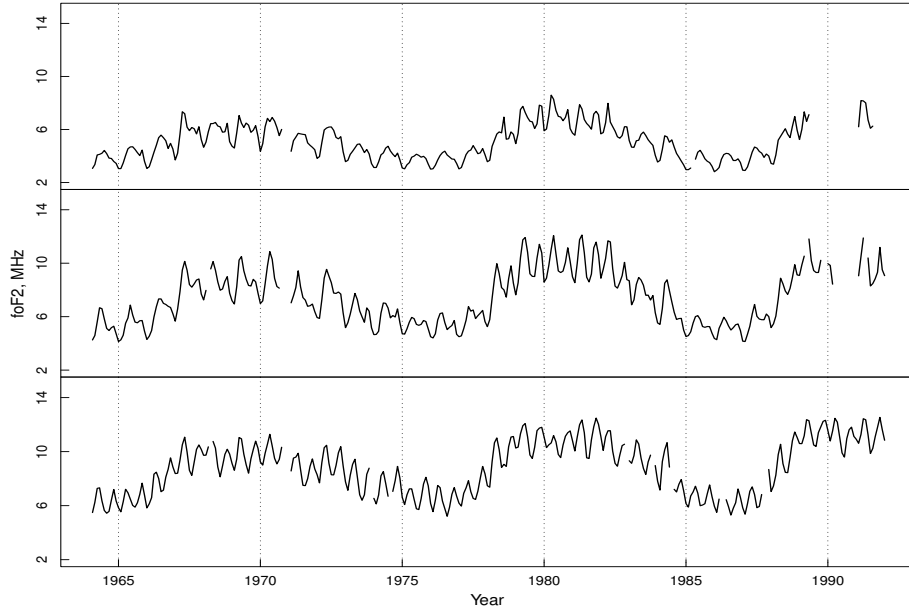
†Mathematics and Statistics Utah State University 3900 Old Main Hill Logan, Utah 84322-3900, USA. Email: `piotr.kokoszka@usu.edu`

Ramsay and Silverman (2005). The annual curves are available at 35 locations, some of which are quite close, and so the curves look very similar, others are very remote with notably different curves. Ramsay and Silverman (2005) use the functional principal components and the functional linear model as exploratory tools. Another example of this type is the Australian rainfall data set, recently used by Delaigle and Hall (2010), which consists of daily rainfall measurements from 1840 to 1990 at 191 Australian weather stations. Due to the importance of such data structures it is useful to investigate when the commonly used techniques designed for iid functional data retain their consistency for spatially distributed data, and when they fail. We establish conditions for consistency, or lack thereof, for the functional mean and the functional principal components. Our conditions combine the spatial dependence of the curves $X(\mathbf{s}_k; \cdot\,)$ and the distribution of the data locations $\mathbf{s}_k$. It is hoped that the general framework we propose will be useful in the development of asymptotic arguments for statistical models involving spatial functional data.

An important example of data that fall into our framework are pollution curves: $X(\mathbf{s}_k; t)$ is the concentration of a pollutant at time t at location $\mathbf{s}_k$. Data of this type were studied by Kaiser *et al.* (2002). A functional framework might be convenient because such data are typically available only at sparsely distributed time points t_j which can be different at different locations. Another interesting example are snow water curves measured at several dozen locations in every state over many decades. Such data have been studied in the spatial framework, e.g. Carroll *et al.* (1995) and Carroll and Cressie (1996), but useful insights can be gained by studying the whole curves reflecting the temporal dynamics. In many studies, $X(\mathbf{s}_k; t)$ is the count at time t of disease cases, where $\mathbf{s}_k$ represents an average location in an areal model.

The data set that most directly motivated this research consists of the curves of the so–called F2-layer critical frequency $foF2$. Three such curves are shown in Figure 1.1. In principle, $foF2$ curves are available at close to 200 locations throughout the globe, but sufficiently complete data are available at only 30-40 locations which are very unevenly spread; for example, there is a dense network of observatories over Europe and practically no data over the oceans. The study of this data set has been motivated by the hypothesis of Roble and Dickinson (1989) who suggested that the increasing amounts of greenhouse gases should lead to global cooling in mesosphere and thermosphere, as opposed to the global warming in lower troposphere. Rishbeth (1990) pointed out that such cooling would result in a thermal contraction and the global lowering of the ionospheric peak densities, which can be computed from the critical frequency $foF2$. The last twenty years have seen very extensive research in this area, see Lastovicka *et al.* (2008) for a partial overview. One of the difficulties is in finding a global trend for curves which appear to exhibit trends in opposing directions over various regions. Ulich *et al.* (2003) stressed that to make any trends believable, a suitable statistical modeling, and a proper treatment of "errors

FIGURE 1.1 F2-layer critical frequency curves at three locations. Top to bottom (latitude in parentheses): Yakutsk (62.0), Yamagawa (31.2), Manila (14.7). The functions must be divided by a deterministic function of the latitude to obtain a stationary field.

and uncertainties" is called for. Space physics data measured at terrestrial observatories always come in the form of temporal curves at fixed spatial locations. Maslova *et al.* (2009, 2010a, 2010b) used the tools of functional data analysis to study such data, but the spatial dependence of the curves was not fully exploited.

There has not been much research on fundamental properties of spatially distributed functional data. Delicado *et al.* (2010) review recent contributions to the methodology for spatially distributed functional data. For geostatistical functional data, several exploratory approaches to kriging have been proposed. Typically fixed basis expansions are used, see Yamanishi and Tanaka (2003) and Bel *et al.* (2010). A general theoretical framework has to address several problems. The first issue is the dimensionality of the index space. While in time series analysis, the process is indexed by an equispaced scalar parameter, we need here a *d*-dimensional index space. For model building this makes a big difference since the dynamics and dependence of the process have to be described in all directions, and the typical recurrence equations used in time series cannot be employed. The model building is further complicated by the fact that the index space is often continuous (geostatistical data). Rather than defining a random field $\{\xi(\mathbf{s}); \mathbf{s} \in \mathbb{R}^d\}$ via a specific model equations, dependence conditions are imposed, in terms of the decay of the covariances or using mixing conditions. Another feature peculiar to random field theory is the design of the sampling points; the distances between them play a fundamental role. Different asymptotics hold in the presense of clusters and for sparsely distributed points.

At least three types of point distributions have been considered (Cressie (1993)): When the region R_N where the points $\{\mathbf{s}_{i,N}; 1 \leq i \leq N\}$ are sampled remains bounded, then we are in the so-called *infill domain sampling* case. Classical asymptotic results, like the law of large numbers or the central limit theorem will usually fail, see Lahiri (1996). The other extreme situation is described by the *increasing domain sampling*. Here a minimum separation between the sampling points $\{\mathbf{s}_{i,N}\} \in R_N$ for all i and N is required. This is of course only possible if $\mathrm{diam}(R_N) \to \infty$. We shall also explore the *nearly infill* situation studied by Lahiri (2003) and Park *et al.* (2009). In this case the domain of the sampling region becomes unbounded ($\mathrm{diam}(R_N) \to \infty$), but at the same time the number of sites in any given subregion tends to infinity, i.e. the points become more dense. These issues are also studied by Zhang (2004), Loh (2005), Lahiri and Zhu (2006), Du *et al.* (2009). We formalize these concepts in Sections 3 and 4. Finally, the interplay of the geostatistical spatial structure and the functional temporal structure must be cast into a workable framework.

For the reasons explained above, the approach of Hörmann and Kokoszka (2010), who developed a framework for estimation and testing for functional time series is totally inappropriate for functional spatial fields. The starting point for the theory of Hörmann and Kokoszka (2010) is the representation $X_k = f(\varepsilon_k, \varepsilon_{k-1}, \ldots)$ of a function X_k in terms of iid error functions ε_k. While all time series models used in practice admit such a representation, no analog representations exist for geostatistical spatial data. (Even though not widely used, spatial autoregressive processes have been proposed, but no Volterra type expansions have been developed for them.)

The paper is organized as follows. Section 2 describes in greater detail the objectives of this research by developing several examples which show how spatially distributed functional data differ from functional random samples and from functional time series. In simple settings, it illustrates what kind of consistency or inconsistency results can be expected, and what kind of difficulties must be overcome. Assumptions on the functional random fields we study are introduced in Section 3. A crucial part of these assumptions consists of conditions on the spatial distribution of the points $\mathbf{s}_k$. Section 4 compares our conditions to those typically assumed for scalar spatial processes. In Sections 5 and 6 we establish consistency results, respectively, for the functional mean and the covariance operator. These sections also contain examples specializing the general results to more specific settings. Section 7 explains, by means of general theorems and examples, when the sample principal components are not consistent. The proofs of the main results are collected in Section 8.

2 Motivating examples

Functional principal components play a fundamental role in functional data analysis, much greater than the usual multivariate principal components. This is mostly due to the fact that the Karhunen-Loève expansion allows to represent functional data in a concise way. This property has been extensively used and studied in various settings. To name only a few illustrative references, we cite Yao *et al.* (2005) Hall and Hosseini-Nasab (2006), Reiss and Ogden (2007), Gabrys and Kokoszka (2007), Benko *et al.* (2009), Paul and Peng (2009), Jiang and Wang (2010) and Gabrys *et al.* (2010). Depending on the structure of the data, theoretical analyses emphasize various aspects of the estimation process, with smoothing in iid samples having being particularly carefully studied. This paper focuses on the spatial dependence and distribution of the curves, which has received no attention so far.

Suppose $X_1, X_2, \ldots, X_N$ are mean zero identically distributed elements of $L^2 = L^2([0,1])$ such that $E\|X\|^4 < \infty$, where the norm is the usual norm generated by the inner product in L^2. The covariance operator is then defined for $x \in L^2$ by $C(x) = E[\langle X, x \rangle X]$. Its eigenfunctions are the functional principal components (FPC's), denoted v_k. Up to a sign, they are estimated by the empirical FPC's (EFPC's), denoted $\hat{v}_k$ and defined as the eigenfunctions of the empirical covariance operator

$$\widehat{C}_N(x) = \frac{1}{N} \sum_{n=1}^{N} \langle X_n, x \rangle X_n, \quad x \in L^2.$$

The distance between v_k and $\hat{v}_k$ is determined by the distance between C and $\widehat{C}_N$. This follows from Lemma 2.1, which has been often used. To state it, consider two compact operators C and K with singular value decompositions

(2.1)
$$C(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle f_j, \quad K(x) = \sum_{j=1}^{\infty} \gamma_j \langle x, u_j \rangle g_j.$$

Recall that a linear operator K in a separable Hilbert space H is said to be Hilbert-Schmidt, if for some orthonormal basis $\{e_i\}$ of H

$$\|K\|_{\mathcal{S}}^2 := \sum_{i \geq 1} \|K(e_i)\|_H^2 < \infty.$$

Then $\|\cdot\|_{\mathcal{S}}$ defines a norm on the space of all operators satisfying this condition. The norm is independent of the choice of the basis. This space is again a Hilbert space with the inner product

$$\langle K_1, K_2 \rangle_{\mathcal{S}} = \sum_{i \geq 1} \langle K_1(e_i), K_2(e_i) \rangle.$$

Set

$$v_j' = \hat{c}_j v_j, \quad \hat{c}_j = \text{sign}(\langle u_j, v_j \rangle).$$

5

LEMMA 2.1 *Suppose C, K are two compact operators with singular value decompositions (2.1). If C and K are Hilbert–Schmidt, symmetric and positive definite, and the eigenvalues of C satisfy*

(2.2) $$\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1},$$

then

$$\|u_j - v'_j\| \leq \frac{2\sqrt{2}}{\alpha_j}\|K - C\|_{\mathcal{S}}, \quad 1 \leq j \leq d,$$

where $\alpha_1 = \lambda_1 - \lambda_2$ and $\alpha_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$, $2 \leq j \leq d$.

Lemma 2.1 can be proven using Corollary 1.6 on p. 99 of Gohberg *et al.* (1990) and following the lines of the proof of Lemma 4.3 of Bosq (2000).

If the functional observations $X_k, k \in \mathbb{Z}$, are independent, then

(2.3) $$\limsup_{N \to \infty} NE\|\widehat{C}_N - C\|_{\mathcal{S}}^2 < \infty.$$

Consequently, for such functional observations, under (2.2),

$$\max_{1 \leq k \leq d} E\|\hat{c}_k \hat{v}_k - v_k\|^2 = O\left(N^{-1}\right).$$

Hörmann and Kokoszka (2010) showed that (2.3) continues to hold for weakly dependent time series, in particular for m–dependent X_k. Our first example shows why m–dependence does not imply (2.3) for spatially distributed data.

EXAMPLE 2.1 Suppose $X_k = X(\mathbf{s}_k)$, where $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$ are points in an arbitrary metric space, and the random field $X(\cdot)$ is such that $X(\mathbf{s})$ is independent of $X(\mathbf{s}')$ if the distance between $\mathbf{s}$ and $\mathbf{s}'$, $d(\mathbf{s}, \mathbf{s}')$, is greater than m. (We continue to assume that the X_k have the same distribution.) Set

$$B_N(m) = \{(k, \ell) : 1 \leq k, \ell \leq N \text{ and } d(\mathbf{s}_k, \mathbf{s}_\ell) \leq m\},$$

and denote by $|B_N(m)|$ the count of pairs in $B_N(m)$. A brief calculation, which uses the Cauchy inequality twice, leads to the bound

$$NE\|\widehat{C}_N - C\|_{\mathcal{S}}^2 \leq N^{-1}|B_N(m)|E\|X(\mathbf{s})\|^4.$$

If the $\mathbf{s}_k$ are the points in $\mathbb{R}^d$ with integer coordinates, then $|B_N(m)|$ is asymptotically proportional to mN, implying $\limsup_{N \to \infty} N^{-1}|B_N(m)| < \infty$, and the standard rate (2.3). But if there are too many pairs in $B_N(m)$ this rate will no longer hold.

Example 2.1 shows that if the points $\mathbf{s}_k$ are not equispaced and too densely distributed, then the standard rate (2.3) will no longer hold. The next example shows that in such cases the EFPC's $\hat{v}_k$ may not converge at all.

EXAMPLE 2.2 Consider a functional random field

$$(2.4) \qquad X(\mathbf{s}; t) = \sum_{j=1}^{\infty} \xi_j(\mathbf{s}) e_j(t), \quad \mathbf{s} \in \mathbb{R}^d, \ t \in [0, 1],$$

where $\{e_j, j \geq 1\}$ is a complete orthonormal system and the $\xi_j(\mathbf{s})$ are mean zero random variables with $E[\xi_j(\mathbf{s}_1)\xi_i(\mathbf{s}_2)] = 0$ if $i \neq j$ and $E[\xi_j(\mathbf{s})\xi_j(\mathbf{s} + \mathbf{h})] = \lambda_j \rho_j(h), h = \|\mathbf{h}\|$, where $\sum_{j=1}^{\infty} \lambda_j < \infty$ and each $\rho_j(\cdot)$ is a positive correlation function. Direct verification shows that $C(x) = \sum_{j=1}^{\infty} \lambda_j \langle e_j, x \rangle e_j$, so the λ_j are the eigenvalues of C, and the e_j the corresponding eigenfunctions.

Now consider a sequence $\mathbf{s}_n \to \mathbf{0}$. Because of the positive dependence, $X(\mathbf{s}_n)$ is close to $X(\mathbf{0})$, so $\widehat{C}_N$ is close to the random operator $X^\star = \langle X(\mathbf{0}), \cdot \rangle X(\mathbf{0})$. Observe that $X^\star(X(\mathbf{0})) = \|X(\mathbf{0})\|^2 X(\mathbf{0})$. Thus $\|X(\mathbf{0})\|^2 = \sum_{j=1}^{\infty} \xi_j^2(\mathbf{0})$ is an eigenvalue of $X^\star$. Since it is random, it cannot be close to any of the λ_j. The eigenfunctions of $\widehat{C}_N$ are also close to random functions in L^2, and do not converge to the FPC's e_j.

The intuition presented in this example is formalized in Section 7, where a specific numerical example is also given.

The above example shows that if the points $\mathbf{s}_n$ are too close to each other, then the empirical functional principal components are not consistent estimates of the population principal components. Other examples of the lack of consistency are known, see Johnstone and Lu (2009) and references therein. They fall into the "small n large p" framework, and the lack of consistency is due to noisy data which are not sparsely represented. A solution is to perform the principal component analysis on transformed data which admits a sparse representation. The spatial functional data that motivate this research admit a natural sparse representation, the lack of consistency is due to dependence and densely distributed locations of the observations. It is not crucial that the $\mathbf{s}_n$ be close to each other. What matters is the interplay of the spatial distances between these points and the strength of dependence between the curves. To illustrate, suppose in Example 2.2, the covariance between $X(\mathbf{s}_n)$ and $X(\mathbf{0})$ is

$$E[\langle X(\mathbf{s}_n), x \rangle \langle X(\mathbf{0}), y \rangle] = \sum_{j=0}^{\infty} \lambda_j \exp\left\{ -\frac{\|\mathbf{s}_n\|}{\rho_j} \right\} \langle e_j, x \rangle \langle e_j, y \rangle.$$

In a finite sample, small $\|\mathbf{s}_n\|$ have the same effect as large ρ_j, i.e. as stronger dependence.

These considerations show that it is useful to have general criteria for functional spatial data, which combine the spatial distribution of the points and the strength of dependence, and which ensure that the functional principal components can be consistently estimated, and, consequently, that further statistical inference for spatial functional data can be carried out. Such criteria should hold for practically useful models for functional spatial

data. The next example discusses such models, with a rigorous formulation presented in Section 3.

EXAMPLE 2.3 Suppose $\{e_j, j \geq 1\}$ is an arbitrary *fixed* orthonormal basis in L^2. Under very mild assumptions, every constant mean functional random field admits the representation

$$(2.5) \qquad X(\mathbf{s}) = \mu + \sum_{j \geq 1} \xi_j(\mathbf{s}) e_j,$$

where the $\xi_j(\mathbf{s})$ are zero mean random variables. In principle, all properties of X, including the spatial dependence structure, can be equivalently stated as properties of the family of the scalar fields ξ_j. Representation (2.5) is thus the most natural and convenient model for spatially distributed functional data. More generally, the mean μ may itself depend on the spatial location $\mathbf{s}$, but in this paper we study spatially stationary random fields.

Assume that $\mu = 0$ and the field X is strictly stationary (in space); see Section 3 for a definition. Suppose we want to predict $X(\mathbf{s}_0)$ using a linear combination of the curves $X(\mathbf{s}_1), X(\mathbf{s}_2), \ldots, X(\mathbf{s}_N)$, i.e. we want to minimize

$$(2.6) \qquad E \left\| X(\mathbf{s}_0) - \sum_{n=1}^{N} a_n X(\mathbf{s}_n) \right\|^2$$

$$= E \langle X(\mathbf{s}_0), X(\mathbf{s}_0) \rangle - 2 \sum_{n=1}^{N} a_n E \langle X(\mathbf{s}_n), X(\mathbf{s}_0) \rangle + \sum_{k,\ell=1}^{N} a_k a_\ell E \langle X(\mathbf{s}_k), X(\mathbf{s}_\ell) \rangle.$$

Thus for the problem of the least squares linear prediction of a mean zero spatial process we need to know only

$$(2.7) \qquad K(\mathbf{s}, \mathbf{s}') = E \left[\langle X(\mathbf{s}), X(\mathbf{s}') \rangle \right].$$

By the orthonormality of the e_j in (2.4),

$$E \left[\langle X(\mathbf{s}), X(\mathbf{s}') \rangle \right] = E \left[\left\langle \sum_{j=1}^{\infty} \xi_j(\mathbf{s}) e_j, \sum_{i=1}^{\infty} \xi_i(\mathbf{s}') e_i \right\rangle \right]$$

$$= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} E[\xi_j(\mathbf{s}) \xi_i(\mathbf{s}')] \langle e_j, e_i \rangle = \sum_{j=1}^{\infty} E[\xi_j(\mathbf{s}) \xi_j(\mathbf{s}')].$$

Thus, the functional covariances (2.7) are fully determined by the covariances

$$(2.8) \qquad K_j(\mathbf{s}, \mathbf{s}') = E[\xi_j(\mathbf{s}) \xi_j(\mathbf{s}')].$$

Notice that we do not need to know the cross covariances $E[\xi_j(\mathbf{s}) \xi_i(\mathbf{s}')]$ for $i \neq j$. Thus, if we are interested in kriging, we can assume that the spatial processes $\xi_j(\cdot)$ in

8

(2.4) are independent. Such an assumption simplifies the verification of some fourth order properties discussed in the following sections. This observation remains true if the spatial field does not have zero mean, i.e. if we observe realizations of $Z(\mathbf{s}) = \mu(\mathbf{s}) + X(\mathbf{s})$. A brief calculation shows that for kriging, it is enough to know $\mu(\cdot)$ and the covariances (2.8). Stein (1999) and Cressie (1993) provide rigorous accounts of kriging for scalar spatial data.

Our next example shows how representation (2.5) and the independence of the ξ_j allow to derive the standard rate (2.3), if the points $\mathbf{s}_k$ are equispaced on the line and the covariances decay exponentially. In the following sections, we construct a theory that allows us to obtain the standard and nonstandard rates of consistency in much more general settings. We will use the following well–known Lemma.

LEMMA 2.2 *Suppose X and Y are jointly normal mean zero random variables such that $EX^2 = \sigma^2$, $EY^2 = \nu^2$, $E[XY] = \rho\sigma\nu$. Then*

$$\mathrm{Cov}(X^2, Y^2) = 2\rho^2\sigma^2\nu^2.$$

EXAMPLE 2.4 Suppose $X(\mathbf{s}; t)$ is an arbitrary functional random field observed at locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$. Then

$$(2.9) \qquad NE\|\widehat{C} - C\|_{\mathcal{S}}^2 = N^{-1} \sum_{k,\ell=1}^{\infty} \iint \mathrm{Cov}(X(\mathbf{s}_k; t)X(\mathbf{s}_k; u), \; X(\mathbf{s}_\ell; t)X(\mathbf{s}_\ell; u))dtdu.$$

Without any further assumptions, a sufficient condition for the EFPC's to be consistent with the rate $N^{-1/2}$ is that the right–hand side of (2.9) is bounded from above by a constant. Under additional assumptions, more precise sufficient conditions are possible.

Suppose first that representation (2.4) holds with independent strictly stationary scalar fields $\xi_j(\cdot)$. Define the covariances

$$E[\xi_j(\mathbf{s}_k)\xi_j(\mathbf{s}_\ell)] = \gamma_j(\mathbf{s}_k - \mathbf{s}_\ell), \quad \mathrm{Cov}(\xi_j^2(\mathbf{s}_k), \xi_j^2(\mathbf{s}_\ell)) = \tau_j(\mathbf{s}_k - \mathbf{s}_\ell).$$

Using (2.9), we see that under these assumptions,

$$NE\|\widehat{C} - C\|_{\mathcal{S}}^2 = N^{-1} \sum_{k,\ell=1}^{\infty} \left\{ \sum_{i \neq j} \gamma_i(\mathbf{s}_k - \mathbf{s}_\ell)\gamma_j(\mathbf{s}_k - \mathbf{s}_\ell) + \sum_{j=1}^{\infty} \tau_j(\mathbf{s}_k - \mathbf{s}_\ell) \right\}.$$

Thus (2.3) holds, if

$$(2.10) \qquad \limsup_{N \to \infty} N^{-1} \sum_{k,\ell=1}^{N} \left\{ \sum_{j=1}^{\infty} \gamma_j(\mathbf{s}_k - \mathbf{s}_\ell) \right\}^2 < \infty$$

9

and

(2.11)
$$\limsup_{N\to\infty} N^{-1} \sum_{k,\ell=1}^{N} \sum_{j=1}^{\infty} |\tau_j(\mathbf{s}_k - \mathbf{s}_\ell)| < \infty.$$

Suppose now, in addition, that X is Gaussian with

(2.12)
$$E[\xi_j(\mathbf{s}_k)\xi_j(\mathbf{s}_\ell)] = \sigma_j^2 \exp\left\{-\rho_j^{-1} d(\mathbf{s}_k, \mathbf{s}_\ell)\right\}$$

so that

(2.13)
$$\mathrm{Cov}(\xi_j^2(\mathbf{s}_k), \xi_j^2(\mathbf{s}_\ell)) = 2\sigma_j^4 \exp\left\{-2\rho_j^{-1} d(\mathbf{s}_k, \mathbf{s}_\ell)\right\}.$$

Suppose the points $\mathbf{s}_k$ are equispaced on the line. Denoting the smallest distance between the points by d, we see that

$$N^{-1} \sum_{k,\ell=1}^{N} \left\{\sum_{j=1}^{\infty} \gamma_j(\mathbf{s}_k - \mathbf{s}_\ell)\right\}^2 = \sum_{j=1}^{\infty} \sigma_j^2 + 2N^{-1} \sum_{m=1}^{N-1} (N-m) \left\{\sum_{j=1}^{\infty} \sigma_j^2 \exp\left(-\rho_j^{-1} m d\right)\right\}^2.$$

If we assume that

(2.14)
$$\sum_{j\geq 1} \sigma_j^2 < \infty \quad \text{and} \quad \sup_{j\geq 1} \rho_j < \rho < \infty,$$

then Conditions (2.10) and (2.11) hold. Condition (2.14) means that the correlation functions of all processes $\xi_j(\cdot)$ must decay uniformly sufficiently fast.

To verify (2.10), observe that

$$N^{-1} \sum_{m=1}^{N-1} (N-m) \left\{\sum_{j=1}^{\infty} \sigma_j^2 \exp\left(-\rho_j^{-1} m d\right)\right\}^2$$

$$\leq \sum_{m=1}^{N-1} \left\{\sum_{j=1}^{\infty} \sigma_j^2 \exp\left(-\rho_j^{-1} m d\right)\right\}^2 \leq \sum_{m=1}^{N-1} \left\{\sum_{j=1}^{\infty} \sigma_j^2 \exp\left(-\rho^{-1} m d\right)\right\}^2$$

$$= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \sigma_j^2 \sigma_i^2 \sum_{m=1}^{N-1} \exp\left(-2\rho^{-1} m d\right) = O(1) \left(\sum_{j=1}^{\infty} \sigma_j^2\right)^2 = O(1).$$

The verification of (2.11) is analogous because (2.14) implies $\sum_{j=1}^{\infty} \sigma_j^4 < \infty$.

We will see that Condition (2.14) (formulated analogously for several classes of models) is applicable in much more general settings than equispaced points on the line.

10

3 Models and Assumptions

We assume $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is a random field taking values in $L^2 = L^2([0,1])$, i.e. each $X(\mathbf{s})$ is a square integrable function defined on $[0,1]$. The value of this function at $t \in [0,1]$ is denoted by $X(\mathbf{s}; t)$. With the usual inner product in L^2, the norm of $X(\mathbf{s})$ is

$$\|X(\mathbf{s})\| = \left\{ \int X^2(\mathbf{s}; t)dt \right\}^{1/2}.$$

We assume that the spatial process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is strictly stationary, i.e. for every $\mathbf{h} \in \mathbb{R}^d$,

$$(3.1) \qquad \big(X(\mathbf{s}_1), X(\mathbf{s}_2), \ldots, X(\mathbf{s}_k) \big) \overset{d}{=} \big(X(\mathbf{s}_1 + \mathbf{h}), X(\mathbf{s}_2 + \mathbf{h}), \ldots, X(\mathbf{s}_k + \mathbf{h}) \big).$$

We also assume that it is square integrable in the sense that

$$(3.2) \qquad\qquad\qquad E\|X(\mathbf{s})\|^2 < \infty.$$

Under (3.1) and (3.2), the common mean function is denoted by $\mu = EX(\mathbf{s})$.

The first question is how we can assure the existence of stationary spatial functional models. A most direct and convenient way is to directly construct them by using (2.5). As $\{e_j\}$ is a basis, every a priori given functional field X admits expansion (2.5). Since $\xi_j(\mathbf{s}) = \langle X(\mathbf{s}) - \mu, e_j \rangle$, the functional field X is strictly stationary if and only if each scalar field ξ_j is strictly stationary. By Parseval's identity $\|X(\mathbf{s}) - \mu\|^2 = \sum_{j\geq 1} \xi_j^2(\mathbf{s})$, so (3.2) holds if and only if $\sum_{j\geq 1} E\xi_j^2(\mathbf{s}) < \infty$.

The cross–covariance operators are defined by

$$C_{\mathbf{s}_1, \mathbf{s}_2}(x) := E\left[\langle X(\mathbf{s}_1) - \mu, x \rangle (X(\mathbf{s}_2) - \mu) \right]$$
$$(3.3) \qquad\qquad = \sum_{k\geq 1} \sum_{j\geq 1} E\big[\xi_j(\mathbf{s}_1)\xi_k(\mathbf{s}_2) \big] \langle e_j, x \rangle e_k.$$

In particular, the covariance operator C is defined by

$$(3.4) \qquad C(x) = C_{\mathbf{s},\mathbf{s}}(x) = E\left[\langle X(\mathbf{s}) - \mu, x \rangle (X(\mathbf{s}) - \mu) \right].$$

If a process has the representation (2.5) with uncorrelated random fields $\xi_j(\cdot)$, i.e.

$$E\big[\xi_i(\mathbf{s}_1)\xi_j(\mathbf{s}_2) \big] = 0 \quad \text{for all } \mathbf{s}_1, \mathbf{s}_2 \text{ if } i \neq j,$$

then the e_j are the eigenfunctions of C and the $\lambda_j = E\xi_j^2(\mathbf{s})$ are the corresponding eigenvalues.

To develop an estimation framework, we impose conditions on the decay of the cross–covariances $E[\langle X(\mathbf{s}_1) - \mu, X(\mathbf{s}_2) - \mu \rangle]$, as the distance between $\mathbf{s}_1$ and $\mathbf{s}_2$ increases. We shall use the distance function defined by the Euclidian norm in $\mathbb{R}^d$, denoted $\|\mathbf{s}_1 - \mathbf{s}_2\|_2$, but other distance functions can be used as well.

ASSUMPTION 3.1 *The spatial process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is strictly stationary and square integrable, i.e. (3.1) and (3.2) hold. In addition,*

$$(3.5) \qquad |E\langle X(\mathbf{s}_1) - \mu, X(\mathbf{s}_2) - \mu\rangle| \le h(\|\mathbf{s}_1 - \mathbf{s}_2\|_2),$$

where $h : [0, \infty) \to [0, \infty)$ with $h(x) \searrow 0$, as $x \to \infty$.

If the process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ has representation (2.5) with some basis $\{e_j\}$, then it can be easily seen from (3.3) that (3.5) is equivalent to

$$(3.6) \qquad \left| \sum_{j \ge 1} \langle C_{\mathbf{s}_1, \mathbf{s}_2}(e_j), e_j\rangle \right| \le h(\|\mathbf{s}_1 - \mathbf{s}_2\|_2).$$

Notice also the relation

$$\langle C_{\mathbf{s}_1, \mathbf{s}_2}(e_j), e_j\rangle = E\big[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)\big],$$

which follows also from (3.3). If we assume more specifically that

$$(3.7) \qquad E\big[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)\big] = \phi_j(\|\mathbf{s}_1 - \mathbf{s}_2\|_2),$$

then (3.5) is equivalent to

$$(3.8) \qquad \left| \sum_{j \ge 1} \phi_j\big(\|\mathbf{s}_1 - \mathbf{s}_2\|_2\big) \right| \le h(\|\mathbf{s}_1 - \mathbf{s}_2\|_2).$$

Examples 3.1 and 3.2 consider typical spatial covariance functions, and show when condition (3.8) holds with a function h as in Assumption 3.1.

EXAMPLE 3.1 Suppose that the fields $\{\xi_j(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$, $j \ge 1$, are zero mean, strictly stationary and α-*mixing.* That is

$$\sup_{(A,B) \in \sigma(\xi_j(\mathbf{s})) \times \sigma(\xi_j(\mathbf{s}+\mathbf{h}))} |P(A)P(B) - P(A \cap B)| \le \alpha_j(\mathbf{h}),$$

with $\alpha_j(\mathbf{h}) \to 0$ if $\|\mathbf{h}\|_2 \to \infty$. Let $\alpha_j'(h) = \sup\{\alpha_j(\mathbf{h}) : \|\mathbf{h}\|_2 = h\}$. Then $\alpha_j^*(h) = \sup\{\alpha_j'(x) : x \ge h\} \searrow 0$ as $h \to \infty$. Using stationarity and the main result in Rio (1993) it follows that

$$|E[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)]| = |E[\xi_j(\mathbf{0})\xi_j(\mathbf{s}_2 - \mathbf{s}_1)]|$$
$$\le 2 \int_0^{2\alpha_j(\mathbf{s}_2 - \mathbf{s}_1)} Q_j^2(u)du$$
$$\le 2 \int_0^{2\alpha_j^*(\|\mathbf{s}_2 - \mathbf{s}_1\|_2)} Q_j^2(u)du$$
$$=: \phi_j(\|\mathbf{s}_2 - \mathbf{s}_1\|_2),$$

12

where $Q_j(u) = \inf\{t : P(|\xi_j(\mathbf{0})| > t) \leq u\}$ is the quantile function of $|\xi_j(\mathbf{0})|$. Note that $\alpha_h(\mathbf{h}) \leq 1/4$ for any $\mathbf{h}$, and thus $\phi_j(x) \leq 2\int_0^1 Q_j^2(u)du = 2E[\xi_j^2(\mathbf{0})]$. If $\sum_{j\geq 1} E\xi_j^2(\mathbf{0}) < \infty$, then (3.5) holds with $h(x) = \sum_{j\geq 1} \phi_j(x)$. (Note that $|h(x)| \searrow 0$ follows from $\alpha_j^*(x) \searrow 0$ and the monotone convergence theorem.)

EXAMPLE 3.2 Suppose (3.7) holds, and set $h(x) = \sum_{j\geq 1} \phi_j(x)$. If each ϕ_j is a powered exponential covariance function defined by

$$\phi_j(x) = \sigma_j^2 \exp\left\{-\left(\frac{x}{\rho_j}\right)^p\right\}.$$

then h satisfies the conditions of Assumption 3.1 if

(3.9)
$$\sum_{j\geq 1} \sigma_j^2 < \infty \quad \text{and} \quad \sup_{j\geq 1} \rho_j < \infty.$$

Condition (3.9) is also sufficient if all ϕ_j are in the Matérn class, see Stein (1999), with the same ν, i.e.

$$\phi_j(x) = \sigma_j^2 x^\nu K_\nu(x/\rho_j),$$

because the modified Bessel function K_ν decays monotonically and approximately exponentially fast; numerical calculations show that $K_\nu(s)$ practically vanishes if $s > \nu$. Condition (3.9) is clearly sufficient for spherical ϕ_j defined (for $d = 3$) by

$$\phi_j(x) = \begin{cases} \sigma_j^2\left(1 - \frac{3x}{2\rho_j} + \frac{x^3}{2\rho_j^3}\right), & x \leq \rho_j \\ 0, & x > \rho_j \end{cases}$$

because ϕ_j is decreasing on $[0, \rho_j]$.

Assumption 3.1 is appropriate when studying estimation of the mean function. For the estimation of the covariance operator, we need to impose a different assumption. Recall that if z and y are elements in some Hilbert space H with norm $\|\cdot\|_H$, the operator $z \otimes y$, is defined by $z \otimes y(x) = \langle z, x\rangle y$. In the following assumption, we suppose that the mean of the functional field is zero. This is justified by notational convenience and because we deal with the consistent estimation of the mean function separately.

ASSUMPTION 3.2 *The spatial process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is strictly stationary with zero mean and with 4 moments, i.e. $E\langle X(\mathbf{s}), x\rangle = 0$, $\forall x \in L^2$, and $E\|X(\mathbf{s})\|^4 < \infty$. In addition,*

(3.10)
$$\left| E\langle X(\mathbf{s}_1) \otimes X(\mathbf{s}_1) - C, X(\mathbf{s}_2) \otimes X(\mathbf{s}_2) - C\rangle_{\mathcal{S}}\right| \leq H\left(\|\mathbf{s}_1 - \mathbf{s}_2\|_2\right),$$

where $H : [0, \infty) \to [0, \infty)$ with $H(x) \searrow 0$, as $x \to \infty$.

Assumption 3.2 cannot be verified using only conditions on the covariances of the scalar fields ξ_j in (2.5) because these covariances do not specify the 4th order structure of the model. This can be done if the random field is Gaussian, as illustrated in Example 6.1, or if additional structure is imposed. If the scalar fields $\xi_i(\cdot)$ are independent, the following Lemma can be used to verify (3.10).

LEMMA 3.1 *Let $X(\mathbf{s})$ have representation (2.5) with zero mean and $E\|X(\mathbf{s})\|^4 < \infty$. Assume further that $\xi_i(\cdot)$ and $\xi_j(\cdot)$ are independent if $i \neq j$. Then*

$$\left| E\langle X(\mathbf{s}_1) \otimes X(\mathbf{s}_1) - C \,, X(\mathbf{s}_2) \otimes X(\mathbf{s}_2) - C \rangle_{\mathcal{S}} \right|$$
$$\leq \left| \sum_{j \geq 1} \mathrm{Cov}\big(\xi_j^2(\mathbf{s}_1), \xi_j^2(\mathbf{s}_2)\big) \right| + \left| \sum_{j \geq 1} E\big[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)\big] \right|^2.$$

PROOF: If $\xi_i(\cdot)$ and $\xi_j(\cdot)$ are independent for $i \neq j$, then the e_j are the eigenvalues of C, and the $\xi_j(\mathbf{s})$ are the principal component scores with $E\xi_j^2(\mathbf{s}) = \lambda_j$. Using continuity of the inner product and dominated convergence we obtain

$$\left| E\langle X(\mathbf{s}_1) \otimes X(\mathbf{s}_1) - C \,, X(\mathbf{s}_2) \otimes X(\mathbf{s}_2) - C \rangle_{\mathcal{S}} \right|$$
$$= \left| E \sum_{j \geq 1} \Big\langle \langle X(\mathbf{s}_1), e_j\rangle X(\mathbf{s}_1) - C(e_j) \,, \langle X(\mathbf{s}_2), e_j\rangle X(\mathbf{s}_2) - C(e_j) \Big\rangle \right|$$
$$= \left| E \sum_{j \geq 1} \Big\langle \xi_j(\mathbf{s}_1) \sum_{\ell \geq 1} \xi_\ell(\mathbf{s}_1)e_\ell - \lambda_j e_j \,, \xi_j(\mathbf{s}_2) \sum_{k \geq 1} \xi_k(\mathbf{s}_2)e_k - \lambda_j e_j \Big\rangle \right|$$
$$= \left| E \sum_{j \geq 1} \Big\{ \xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2) \sum_{\ell \geq 1} \xi_\ell(\mathbf{s}_1)\xi_\ell(\mathbf{s}_2) + \lambda_j^2 - \lambda_j \xi_j^2(\mathbf{s}_1) - \lambda_j \xi_j^2(\mathbf{s}_2) \Big\} \right|$$
$$\leq \left| \sum_{j \geq 1} \mathrm{Cov}\big(\xi_j^2(\mathbf{s}_1), \xi_j^2(\mathbf{s}_2)\big) \right| + \left| \sum_{j \geq 1} \sum_{\ell \neq j} E\big[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)\big] \times E\big[\xi_\ell(\mathbf{s}_1)\xi_\ell(\mathbf{s}_2)\big] \right|$$
$$\leq \left| \sum_{j \geq 1} \mathrm{Cov}\big(\xi_j^2(\mathbf{s}_1), \xi_j^2(\mathbf{s}_2)\big) \right| + \left| \sum_{j \geq 1} E\big[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)\big] \right|^2.$$

∎

As already noted, for spatial processes assumptions on the distribution of the sampling points are as important as those on the covariance structure. To formalize the different sampling schemes introduced in Section 1, we propose the following measure of "minimal dispersion" of some point cloud $\mathfrak{S}$:

$$I_\rho(\mathbf{s}, \mathfrak{S}) = |\{\mathbf{y} \in \mathfrak{S} : \|\mathbf{s} - \mathbf{y}\|_2 \leq \rho\}|/|\mathfrak{S}| \quad \text{and} \quad I_\rho(\mathfrak{S}) = \sup\{I_\rho(\mathbf{s}, \mathfrak{S}), \ \mathbf{s} \in \mathfrak{S}\},$$

where $|\mathfrak{S}|$ denotes the number of elements of $\mathfrak{S}$. The quantity $I_\rho(\mathfrak{S})$ is the maximal fraction of $\mathfrak{S}$–points in a ball of radius ρ centered at an element of $\mathfrak{S}$. Notice that $1/|\mathfrak{S}| \leq I_\rho(\mathfrak{S}) \leq 1$. We call $\rho \mapsto I_\rho(\mathfrak{S})$ the *intensity function* of $\mathfrak{S}$.

14

DEFINITION 3.1 For a sampling scheme $\mathfrak{S}_N = \{\mathbf{s}_{i,N}; 1 \le i \le S_N\}$, $S_N \to \infty$, we consider the following conditions:

(i) there is a $\rho > 0$ such that $\limsup_{N \to \infty} I_\rho(\mathfrak{S}_N) > 0$;
(ii) for some sequence $\rho_N \to \infty$ we have $I_{\rho_N}(\mathfrak{S}_N) \to 0$;
(iii) for any fixed $\rho > 0$ we have $S_N I_\rho(\mathfrak{S}_N) \to \infty$.

We call a deterministic sampling scheme $\mathfrak{S}_N = \{\mathbf{s}_{i,N}; 1 \le i \le S_N\}$

Type A if (i) holds;
Type B if (ii) and (iii) hold;
Type C if (ii) holds, but there is a $\rho > 0$ such that $\limsup_{N \to \infty} S_N I_\rho(\mathfrak{S}_N) < \infty$.

If the sampling scheme is stochastic we call it *Type A, B* or *C* if relations (i), (ii) and (iii) hold with $I_\rho(\mathfrak{S}_N)$ replaced by $EI_\rho(\mathfrak{S}_N)$.

Type A sampling is related to purely infill domain sampling which corresponds to $I_\rho(\mathfrak{S}_N) = 1$ for all $N \ge 1$, provided ρ is large enough. However, in contrast to the purely infill domain sampling, it still allows for a non-degenerate asymptotic theory for sparse enough subsamples (in the sense of Type B or C).

EXAMPLE 3.3 Assume that $\mathfrak{S}_N$ are sampling points on the line with $s_{2k} = 1/k$ and $s_{2k+1} = k$, $1 \le k \le N$. Then, for $\rho = 1$, $\lim_{N \to \infty} I_\rho(\mathfrak{S}_N) = 1/2$, so this sampling scheme is of Type A. But the subsample corresponding to odd indices is of Type C.

A brief reflection shows that assumptions (i) and (ii) are mutually exclusive. Combining (ii) and (iii) implies that the points intensify (at least at certain spots) excluding the purely increasing domain sampling. Hence the Type B sampling corresponds to the nearly infill domain sampling. If only (ii) holds, but (iii) does not (Type C sampling) then the sampling scheme corresponds to purely increasing domain sampling.

Our conditions are more general than those proposed so far. Their relation to more specific sampling designs previously used is discussed in Section 4.

4 Regular spatial designs

We continue to assume a spatial design $\mathfrak{S}_N = \{\mathbf{s}_{k,N}, 1 \le k \le S_N\}$. The two special cases we discuss are closely related to those considered by Lahiri (2003). The points are assumed to be on a grid of an increasing size, or to have a density. The results of this section show how our more general assumptions look in these special cases, and provide additional intuition behind the sampling designs formulated in Definition 3.1. They also set a framework for some results of Sections 5 and 6.

4.1 Non-random regular design

Let $\mathcal{Z}(\boldsymbol{\delta})$ be a lattice in $\mathbb{R}^d$ with increments δ_i in the i-th direction. Let $\delta_0 = \min\{\delta_1, \ldots, \delta_d\}$, $\Delta^d = \prod_{i=1}^d \delta_i$ and let $R_N = \alpha_N R_0$, where R_0 is some bounded Riemann measurable Borel-set in $\mathbb{R}^d$ containing the origin. A set is Riemann measurable if its indicator function is Riemann integrable. This condition excludes highly irregular sets R_0. The scaling parameters $\alpha_N > 0$ are assumed to be non-decreasing and will be specified below in Lemma 4.2. We assume without loss of generality that $\mathrm{Vol}(R_0) = 1$, hence $\mathrm{Vol}(R_N) = \alpha_N^d$. Typical examples are $R_0 = \{x \in \mathbb{R}^d : \|x\| \leq z_{1,d}\}$, with $z_{1,d}$ equal to the radius of the d-dimensional sphere with volume 1, or $R_0 = [-1/2, 1/2]^d$. The sampling points $\mathfrak{S}_N$ are defined as $\{\mathbf{s}_{k,N}, 1 \leq k \leq S_N\} = \mathcal{Z}(\eta_N \boldsymbol{\delta}) \cap R_N$, where η_N is chosen such that the sample size $S_N \sim N$. It is intuitively clear that $\mathrm{Vol}(R_N) \approx \eta_N^d \Delta^d S_N$, suggesting

$$(4.1) \qquad \eta_N = \frac{\alpha_N}{\Delta N^{1/d}}.$$

A formal proof that η_N in (4.1) assures $S_N \sim N$ is immediate from the following

LEMMA 4.1 *Let K be a bounded set in $\mathbb{R}^d$, and assume that K is Riemann measurable with $\mathrm{Vol}(K) = 1$. If $\beta_N \to 0$, then*

$$|K \cap \mathcal{Z}(\beta_N \boldsymbol{\delta})| \sim \frac{1}{\Delta^d \beta_N^d}.$$

PROOF: Let $K \subset M_1 \subset M_2$ where M_1 and M_2 are rectangles in $\mathbb{R}^d$ having no intersecting margin (M_1 is an inner subset of M_2). The points $\{x_{i,N}\} = \mathcal{Z}(\beta_N \boldsymbol{\delta}) \cap M_2$ can be seen as the vertices of rectangles $J_{i,N} = x_{i,N} + \{\mathbf{t} \circ \beta_N \boldsymbol{\delta}, \mathbf{t} \in [0,1]^d\}$, where $\circ$ denotes the Hadamard (entrywise) product. For large enough N, the sets $L_{i,N} = J_{i,N} \cap M_1$ define a partition of M_1. Then, by the assumed Riemann measurability,

$$\int_{M_1} I_K(x)dx$$
$$= \liminf_{N \to \infty} \beta_N^d \Delta^d \sum_i \inf\{I_K(x) : x \in L_{i,N}\}$$
$$\leq \liminf_{N \to \infty} \beta_N^d \Delta^d \sum_i I_K(x_{i,N})$$
$$\leq \limsup_{N \to \infty} \beta_N^d \Delta^d \sum_i I_K(x_{i,N})$$
$$\leq \limsup_{N \to \infty} \beta_N^d \Delta^d \sum_i \sup\{I_K(x) : x \in L_{i,N}\}$$
$$= \int_{M_1} I_K(x)dx.$$

The following Lemma relates the non-random regular design to Definition 3.1. We write $a_N \gg b_N$ if $\limsup b_N/a_N < \infty$.

LEMMA 4.2 *In the above described design the following pairs of statements are equivalent:*

(i) α_N remains bounded $\Leftrightarrow$ Type A sampling;
(ii) $\alpha_N \to \infty$ and $\alpha_N = o(N^{1/d}) \Leftrightarrow$ Type B sampling;
(iii) $\alpha_N \gg N^{1/d} \Leftrightarrow$ Type C sampling.

PROOF: Let $U_\varepsilon(x)$ be the sphere in $\mathbb{R}^d$ with center x and radius ε. Assume first that $\alpha_N = o(N^{1/d})$, which covers (i) and (ii). In this case the volume of the rectangles $L_{i,n}$ as described in the proof of Lemma 4.1 satisfies

$$(4.2) \qquad \mathrm{Vol}(L_{i,n}) = \Delta^d \eta_N^d = \frac{\alpha_N^d}{N} \to 0.$$

Hence $|U_\rho(x) \cap \mathcal{Z}(\eta_N \boldsymbol{\delta})|$ is asymptotically proportional to

$$\mathrm{Vol}(U_\rho(x))/\mathrm{Vol}(L_{i,n}) = V_d \left(\frac{\rho}{\alpha_N}\right)^d N,$$

where V_d is the volume of the d-dimensional unit sphere. Now if we fix an arbitrary $\rho_0 > 0$ then there are constants $0 < C_L < C_U < \infty$, such that for any $\rho \geq \rho_0$ and $N \geq N_0$ and $x \in \mathbb{R}^d$

$$C_L \left(\frac{\rho}{\alpha_N}\right)^d \leq \frac{|U_\rho(x) \cap \mathcal{Z}(\eta_N \boldsymbol{\delta})|}{N} \leq C_U \left(\frac{\rho}{\alpha_N}\right)^d.$$

By the required Riemann measurability we can find an $x \in R_0$ such that for some small enough ε we have $U_{2\varepsilon}(x) \subset R_0$. Then $U_{2\varepsilon\alpha_N}(\alpha_N x) \subset R_N$. Hence for any $2\rho_0 \leq \rho \leq \varepsilon\alpha_N$,

$$C_L \left(\frac{\rho}{\alpha_N}\right)^d \leq \frac{|U_{\rho/2}(\alpha_N x) \cap \mathfrak{S}_N|}{N} \leq I_\rho(\mathfrak{S}_N) \leq \frac{|U_{2\rho}(\alpha_N x) \cap \mathfrak{S}_N|}{N} \leq C_U \left(\frac{\rho}{\alpha_N}\right)^d.$$

With the help of the above inequalities (i) and (ii) are easily checked.

Now we prove (iii). We notice that by (4.2) $\alpha_N \gg N^{1/d}$ is equivalent to $\mathrm{Vol}(L_{i,n})$ does not converge to 0. Assume first that we have Type C sampling. Then by the arguments above we find an x and a $\rho > 0$ such that $U_\rho(\alpha_N x) \subset R_N$. Thus

$$|U_\rho(\alpha_N x) \cap \mathcal{Z}(\eta_N \boldsymbol{\delta})| \leq S_N I_\rho(\mathfrak{S}_N).$$

As this quantity remains bounded, $\mathrm{Vol}(L_{i,n})$ does not converge to 0.

On the other hand, if $\mathrm{Vol}(L_{i,n})$ does not converge to 0 then for any $\rho > 0$ and any $x \in \mathbb{R}^d$ we have $\limsup_{N\to\infty} |U_\rho(x) \cap \mathcal{Z}(\eta_N \boldsymbol{\delta})| < \infty$ and thus for arbitrary large ρ

$$I_\rho(\mathfrak{S}_N) \leq \sup_x \frac{|U_\rho(x) \cap \mathcal{Z}(\eta_N \boldsymbol{\delta})|}{S_N} \to 0.$$

The claim follows immediately.

4.2   Randomized design

Let $\{\mathbf{s}_k, 1 \leq k \leq N\}$ be iid random vectors with a density $f(\mathbf{s})$ which has support on a Borel set $R_0 \subset \mathbb{R}^d$ containing the origin and satisfying $\mathrm{Vol}(R_0) = 1$. Again we assume Riemann measurability for R_0 to exclude highly irregular sets. For the sake of simplicity we shall assume that on R_0 the density is bounded away from zero, so that we have $0 < f_L \leq \inf_{x \in R_0} f(x)$. The point set $\{\mathbf{s}_{k,N}, 1 \leq k \leq N\}$ is defined by $\mathbf{s}_{k,N} = \alpha_N \mathbf{s}_k$ for $k = 1, \ldots, N$. For fixed N, this is equivalent to: $\{\mathbf{s}_{k,N}, 1 \leq k \leq N\}$ is an iid sequence on $R_N = \alpha_N R_0$ with density $\alpha_N^{-d} f(\alpha_N^{-1}\mathbf{s})$.

We cannot expect to obtain a full analogue of Lemma 4.2 in the randomized setup. For Type C sampling, the problem is much more delicate, and a closer study shows that it is related to the oscillation behavior of multivariate empirical processes. While Stute (1984) gives almost sure upper bounds, we would need here sharp results on the moments of the modulus of continuity of multivariate empirical process. Such results exist, see Einmahl and Ruymgaart (1987), but are connected to technical assumptions on the bandwidth for the modulus (here determined by α_N) which are not satisfied in our setup. Hence a detailed treatment would go beyond the scope of this paper. We thus state here the following lemma.

LEMMA 4.3 *In the above described sampling scheme the following statements hold:*

(i) α_N remains bounded $\Rightarrow$ Type A sampling;
(ii) $\alpha_N \to \infty$ and $\alpha_N = o(N^{1/d}) \Rightarrow$ Type B sampling;

PROOF. By Jensen's inequality we infer that

$$
\begin{aligned}
EI_\rho(\mathfrak{S}_N) = E \sup_{x \in R_N} \frac{1}{N} \sum_{k=1}^{N} I\{\mathbf{s}_{k,N} \in U_\rho(x) \cap R_N\} \\
\geq \sup_{x \in R_N} P\big(\mathbf{s}_{1,N} \in U_\rho(x) \cap R_N\big) \\
= \sup_{x \in R_0} P\big(\mathbf{s}_1 \in U_{\rho/\alpha_N}(x) \cap R_0\big) \\
= \sup_{x \in R_0} \int_{U_{\rho/\alpha_N}(x) \cap R_0} f(s)ds.
\end{aligned}
$$

We have two scenarios. First, α_N remains bounded. Then we can choose ρ big enough such that $U_{\rho/\alpha_N}(0)$ covers R_0 for all N. It follows that $\limsup_{N \to \infty} EI_\rho(\mathfrak{S}_N) = 1$ and (i) follows.

Second, $\alpha_N \to \infty$. Then for large enough N, R_0 contains a ball with radius ρ/α_N. It follows that

$$
(4.3) \qquad\qquad EI_\rho(\mathfrak{S}_N) \geq f_L V_d \left(\frac{\rho}{\alpha_N} \right)^d.
$$

Now statement (ii) follows easily. ■

5 Consistency of the sample mean function

Our goal is to establish the consistency of the sample mean for functional spatial data. We consider Type B or Type C sampling and obtain rates of convergence. We start with a general setup, and show that the rates can be improved in special cases. The general results are applied to functional random fields with specific covariance structures. The proofs of the main results, Propositions 5.1, 5.2,5.3, are collected in Section 8.

For independent or weakly dependent functional observations X_k,

$$(5.1) \qquad E\left\|\frac{1}{N}\sum_{k=1}^{N} X_k - \mu\right\|^2 = O\left(N^{-1}\right).$$

Proposition 5.1 shows that for general functional spatial processes, the rate of consistency may be much slower than $O\left(N^{-1}\right)$; it is the maximum of $h(\rho_N)$ and $I_{\rho_N}(\mathfrak{S}_N)$ with ρ_N from (ii) of Definition 3.1. Intuitively, the sample mean is consistent if there is a sequence of increasing balls which contain a fraction of points which tends to zero, and the decay of the correlations compensates for the increasing radius of these balls.

PROPOSITION 5.1 *Let Assumption 3.1 hold, and assume that $\mathfrak{S}_N$ defines a non-random design of Type A, B or C. Then for any $\rho_N > 0$,*

$$(5.2) \qquad E\left\|\frac{1}{N}\sum_{k=1}^{N} X(\mathbf{s}_{k,N}) - \mu\right\|^2 \le h(\rho_N) + h(0)I_{\rho_N}(\mathfrak{S}_N).$$

Hence, under the Type B or Type C non-random sampling, with ρ_N as in (ii) of Definition 3.1, the sample mean is consistent.

EXAMPLE 5.1 Assume that N points $\{\mathbf{s}_{k,N}, 1 \le k \le N\}$ are on a regular grid in $\alpha_N[-1/2, 1/2]^d$. Then, as we have seen in Section 4.1, $I_\rho(\mathfrak{S}_N)$ is proportional to $(\rho/\alpha_N)^d$.
For example, if $h(x) = 1/(1+x)^2$, then choosing $\rho_N = \alpha_N^{d/(d+2)}$ we obtain that

$$h(\rho_N) + h(0)I_{\rho_N}(\mathfrak{S}_N) \ll \alpha_N^{-2d/(d+2)} \vee N^{-1}.$$

(Recall that $I_{\rho_N}(\mathfrak{S}_N) \ge N^{-1}$.)

A question that needs to be addressed is whether the bound obtained in Proposition 5.1 is optimal. It is not surprising that (5.2) will not be *uniformly* optimal. This is because the assumptions in Proposition 5.1 are too general to give a precise rate for all the cases

covered. For instance, a smaller bound for Example 5.1 is obtained using Proposition 5.2 below. In some sense, however, the rate (5.2) is optimal, as it possible to construct examples which attain the bound (5.2).

EXAMPLE 5.2 Let $X(s;t) = \psi(s)e(t)$, with $s \in \mathbb{R}$, $t \in [0,1]$, $\int_0^1 e^2(t)dt = 1$ and

$$\psi(s) = \sum_{k \in \mathbb{Z}} I\{s \in (U + k, U + k + 1]\}\delta_k,$$

where $\{\delta_k\}$ is an iid sequence with $\delta_1 = \pm 1$, each with probability $1/2$ and U is uniformly distributed on $[0,1]$ and independent of $\{\delta_k\}$. A simple calculation shows that $EX(s;t) = 0$ for all s,t and that $E\langle X(u), X(v)\rangle = (1 - |u - v|)I\{|u - v| \le 1\}$. Let

$$\mathfrak{S}_N = \left\{ \frac{k\alpha_N}{N}, 1 \le k \le N \right\}.$$

This sampling scheme is of Type A, B or C, depending on whether α_N remains bounded, $\alpha_N \to \infty$, $\alpha_N = o(N)$ and $N = O(\alpha_N)$, respectively. In the latter case let us assume for the sake of simplicity that $\alpha_N = N$. Using the explicit formula for $E\langle X(u), X(v)\rangle$ we obtain

$$E \left\| \frac{1}{N} \sum_{k=1}^{N} X(s_k) - \mu \right\|^2 = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \left(1 - |k - \ell|\frac{\alpha_N}{N}\right) I\left\{|k - \ell| \le \frac{N}{\alpha_N}\right\}$$

$$= \frac{1}{N} + \frac{2}{N^2} \sum_{h=1}^{N/\alpha_N} \left(1 - \frac{h\alpha_N}{N}\right)(N - h)$$

$$\sim \begin{cases} N^{-1} & \text{if } \alpha_N = N; \\ \alpha_N^{-1} & \text{if } \alpha_N \to \infty, \ \alpha_N = o(N); \\ 2 \int_0^{\min\{\alpha^{-1},1\}} (1 - \alpha|x|)(1 - |x|)dx & \text{if } \alpha_N \to \alpha. \end{cases}$$

For Type B and Type C sampling, the optimal bound using Proposition 5.1 is obtained setting $\rho_N = 1$, in which case we have that the r.h.s. in (5.2) is $I_1(\mathfrak{S}_N) = N^{-1}$ if $\alpha_N = N$ and $I_1(\mathfrak{S}_N) = \alpha_N^{-1}$ if $\alpha_N = o(N)$. Under Type A sampling the r.h.s. in (5.2) remains bounded away from zero and the same holds true for the exact quadratic loss.

We now consider the special case, where we have a regular sampling design. Here we are able to obtain the strongest results.

PROPOSITION 5.2 *Assume the sampling design of Section 4.1. Let Assumption 3.1 hold with h such that $x^{d-1}h(x)$ is monotone on $[b, \infty)$, $b > 0$. Then under Type B sampling*

$$E \left\| \frac{1}{S_N} \sum_{k=1}^{S_N} X(\mathbf{s}_{k,N}) - \mu \right\|^2$$

(5.3)
$$\le \frac{1}{\alpha_N^d} \left\{ d(3\Delta)^d \int_0^{K\alpha_N} x^{d-1}h(x)dx + o(1) \sup_{x \in [0, K\alpha_N]} x^{d-1}h(x) \right\},$$

20

for some large enough constant K which is independent of N. Under Type C sampling $1/\alpha_N^d$ in (5.3) is replaced by $O(N^{-1})$.

The technical assumptions on h pose no practical problem, they are satisfied for all important examples, see Example 3.2. A common situation is that $x^{d-1}h(x)$ is increasing on $[0, b]$ and decreasing thereafter.

Our first example shows that for most typical covariance functions, under nearly infill domain sampling, the rate of consistency may be much slower than for the iid case, if the size of the domain does not increase fast enough.

EXAMPLE 5.3 Suppose the functional spatial process has representation (2.5), and (3.7) holds with with the covariance functions ϕ_j as in Example 3.2 (powered exponential, Matérn or spherical). Define $h(x) = \sum_{j\geq 1}\phi_j(x)$, and assume that condition (3.9) holds. Assumption 3.1 is then satisfied and

$$(5.4) \qquad \int_0^\infty x^{d-1}h(x)dx < \infty \quad \text{and} \quad \sup_{x\in\mathbb{R}} x^{d-1}h(x) < \infty.$$

Therefore, for the sampling design of Section 4.1,

$$(5.5) \qquad E\left\|\frac{1}{S_N}\sum_{k=1}^{S_N}X(\mathbf{s}_{k,N}) - \mu\right\|^2 = \begin{cases} O\left(\alpha_N^{-d}\vee N^{-1}\right), & \text{under Type B sampling} \\ O\left(N^{-1}\right), & \text{under Type C sampling} \end{cases}$$

The next example shows that formula (5.5) is far from universal, and that the rate of consistency may be even slower if the covariances decay slower than exponential.

EXAMPLE 5.4 Consider the general setting of Example 5.3, but assume that each covariance function ϕ_j has the quadratic rational form

$$\phi_j(x) = \sigma_j^2\left\{1 + \left(\frac{x}{\rho_j}\right)^2\right\}^{-1}.$$

Condition (3.9) implies that $h(x) = \sum_{j\geq 1}\phi_j(x)$ satisfies Assumption 3.1, but now $h(x) \sim x^{-2}$, as $x \to \infty$. Because of this rate, condition (5.4) holds only for $d = 1$ (and so for this dimension (5.5) also holds). If $d \geq 2$, (5.4) fails, and to find the rate of the consistency, we must use (5.3) directly. We focus only on Type B sampling, and assume implicitly that the rate is slower than N^{-1}. We assume (3.9) throughout this example.

If $d = 2$,

$$\int_0^{K\alpha_N} x^{d-1}h(x)dx = \sum_j \sigma_j^2 \int_0^{K\alpha_N} x\left\{1 + \left(\frac{x}{\rho_j}\right)^2\right\}^{-1}dx$$

$$= \sum_j \sigma_j^2\rho_j^2\, O\left(\int_1^{K\alpha_N} x^{-1}dx\right) = O(\ln\alpha_N)$$

21

and similarly $\sup_{x \in [0, K\alpha_N]} x^{d-1} h(x) = O(1)$.

If $d \geq 3$, the leading term is

$$\int_0^{K\alpha_N} x^{d-1} h(x) dx = O\left(\alpha_N^{d-3}\right).$$

We summarize these calculations as

$$E \left\| \frac{1}{S_N} \sum_{k=1}^{S_N} X(\mathbf{s}_{k,N}) - \mu \right\|^2 = \begin{cases} O\left(\alpha_N^{-1}\right), & \text{if } d = 1 \\ O\left(\alpha_N^{-2} \ln(\alpha_N)\right), & \text{if } d = 2 \\ O\left(\alpha_N^{-2}\right), & \text{if } d \geq 3, \end{cases}$$

for Type B sampling scheme (provided the rate is slower than N^{-1}).

The last example shows that for very persistent spatial dependence, the rate of consistency can be essentially arbitrarily slow.

EXAMPLE 5.5 Assume that $h(x)$ decays only at a logarithmic rate, $h(x) = \{\log(x \vee e)\}^{-1}$. Then, for any $d \geq 1$, the left hand side in (5.3) is $\ll (\log \alpha_N)^{-1}$.

We now turn to the case of the random design.

PROPOSITION 5.3 *Assume the random sampling design of Section 4.2. If the sequence* $\{\mathbf{s}_k\}$ *is independent of the process* X, *and if Assumption 3.1 holds, then we have for any* $\varepsilon_N > 0$

$$E \left\| \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_{k,N}) - \mu \right\|^2 \leq V(\varepsilon_N) \sup_{\mathbf{s} \in R_0} f^2(\mathbf{s}) + h(\alpha_N \varepsilon_N) + \frac{h(0)}{N},$$

where

(5.6) $$V(\varepsilon_N) = \mathrm{Vol}\left\{ (\mathbf{s}, \mathbf{r}) \in R_0^2 : \|\mathbf{s} - \mathbf{r}\|_2 \leq \varepsilon_N \right\}.$$

Choosing ε_N *such that* $\varepsilon_N \to 0$ *and* $\alpha_N \varepsilon_N \to \infty$, *it follows that under Type B or Type C sampling, the sample mean is consistent.*

The bound in Proposition 5.3 can be easily applied to any specific random sampling design and any model for the functions ϕ_j in (2.5). It nicely shows that what matters for the rate of consistency is the interplay between between the rate of growth of the sampling domain and the rate of decay of dependence. For typical sets R_0, $V(\varepsilon_N)$ is proportional to ε_N. Taking $\varepsilon_N = N^{-1}$, we see that the rate of consistency is $h(\alpha_N/N) \vee N^{-1}$. For typical covariance functions ϕ_j, like powered exponential, Matérn or spherical, $h(\alpha_N/N)$ decays faster than N^{-1}, provided α_N increases faster than N. In such cases, the rate of consistency is the same as for an iid sample. For ease of reference, we formulate the following corollary, which can be used in practical applications.

COROLLARY 5.1 *Assume the random sampling design of Section 4.2 with the sequence $\{\mathbf{s}_k\}$ independent the process X. Suppose (2.5) and (3.7) hold with the ϕ_j in one of the families specified in Example 3.2. If Condition (3.9) holds, and $\alpha_N \geq aN \ln N$, for some $a > 0$, then (5.1) holds.*

6 Consistency of the empirical covariance operator

In Section 5 we found the rates of consistency for the functional sample mean. We now turn to the rates for the sample covariance operator. Assuming the functional observations have mean zero, the natural estimator of the covariance operator C is the sample covariance operator given by

$$\widehat{C}_N = \frac{1}{N} \sum_{k=1}^{N} X(\mathbf{s}_k) \otimes X(\mathbf{s}_k).$$

In general, the sample covariance operator is defined by

$$\hat{\Gamma}_N = \frac{1}{N} \sum_{k=1}^{N} \left(X(\mathbf{s}_k) - \bar{X}_N \right) \otimes \left(X(\mathbf{s}_k) - \bar{X}_N \right),$$

where

$$\bar{X}_N = \frac{1}{N} \sum_{k=1}^{N} X(\mathbf{s}_k).$$

Both operators are implemented in statistical software packages, for example in the popular R package FDA and in a similar MATLAB package, see Ramsay *et al.* (2009), The operator $\hat{\Gamma}_N$ is used to compute the EFPC's for centered data, while $\widehat{C}_N$ for data without centering.

We first derive the rates of consistency for $\widehat{C}_N$ assuming $EX(\mathbf{s}) = 0$. Then we turn to the operator $\hat{\Gamma}_N$. The proofs are obtained by applying the technique developed for the estimation of the functional mean. It is a general approach based on the estimation of the second moments of an appropriate norm (between estimator and estimand) so that the conditions in Definition 3.1 can come into play. It is broadly applicable to all statistics obtained by simple averaging of quantities defined at single spatial location. The proofs are thus similar to those presented in the simplest case in Section 8, but the notation becomes more cumbersome because of the increased complexity of the objects to be averaged. To conserve space these proofs are not included.

We begin by observing that

$$E\big\|\widehat{C}_N - C\big\|_{\mathcal{S}}^2 = \langle \widehat{C}_N - C \,,\, \widehat{C}_N - C \rangle_{\mathcal{S}}$$

$$= \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \langle X(\mathbf{s}_k) \otimes X(\mathbf{s}_k) - C \,,\, X(\mathbf{s}_\ell) \otimes X(\mathbf{s}_\ell) - C \rangle_{\mathcal{S}}.$$

It follows that under Assumption 3.2

$$(6.1) \qquad E\big\|\widehat{C}_N - C\big\|_{\mathcal{S}}^2 \le \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} H\big(\|\mathbf{s}_k - \mathbf{s}_\ell\|_2\big).$$

Relation (6.1) is used as the starting point of all proofs, cf. the proof of Proposition 5.1 in Section 5. Modifying the proofs of Section 5, we arrive at the following results.

PROPOSITION 6.1 *Let Assumption 3.2 hold, and assume that $\mathfrak{S}_N$ defines a non-random design of Type A, B or C. Then for any $\rho_N > 0$*

$$E\big\|\widehat{C}_N - C\big\|_{\mathcal{S}}^2 \le H(\rho_N) + H(0) I_{\rho_N}(\mathfrak{S}_N).$$

Hence under the Type B or Type C non-random sampling, with ρ_N as in (ii) of Definition 3.1, the empirical covariance operator is consistent.

PROPOSITION 6.2 *Assume the sampling design of Section 4.1. Let Assumption 3.2 hold, with some function H such that $x^{d-1}H(x)$ is monotone on $[b, \infty)$, $b > 0$. Then under Type B sampling*

$$E\big\|\widehat{C}_N - C\big\|_{\mathcal{S}}^2 \le \frac{1}{\alpha_N^d} \left\{ d(3\Delta)^d \int_0^{K\alpha_N} x^{d-1} H(x)\,dx + o(1) \sup_{x \in [0, K\alpha_N]} x^{d-1} H(x) \right\},$$

for some large enough constant K which is independent of N. Under Type C sampling, the factor $1/\alpha_N^d$ is replaced by $O(N^{-1})$.

PROPOSITION 6.3 *Assume the random sampling design of Section 4.2. If the sequence $\{\mathbf{s}_k\}$ is independent the process X and if Assumption 3.2 holds, then we have for any $\varepsilon_N > 0$,*

$$E\left\|\frac{1}{N}\sum_{k=1}^{N} X(\mathbf{s}_{k,N}) - \mu\right\|^2 \le V(\varepsilon_N) \sup_{\mathbf{s} \in R_0} f^2(\mathbf{s}) + H(\alpha_N \varepsilon_N) + \frac{H(0)}{N},$$

with $V(\varepsilon_N)$ given by (5.6).

It follows that under Type B or Type C sampling the sample covariance operator is consistent.

EXAMPLE 6.1 Let X have representation (2.5), in which the scalar fields $\xi_j(\cdot)$ are independent and Gaussian, and (2.12) (2.13) and (2.14) hold.

It follows that for some large enough constant A,

$$\left| \sum_{j \ge 1} \mathrm{Cov}\big(\xi_j^2(\mathbf{s}_1), \xi_j^2(\mathbf{s}_2)\big) \right| + \left| \sum_{j \ge 1} E\big[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)\big] \right|^2$$
$$\le A \exp\big(-2\rho^{-1}\|\mathbf{s}_1 - \mathbf{s}_2\|_2\big).$$

Hence by Lemma 3.1, Assumption 3.2 holds with $H(x) = A \exp\left(-2\rho^{-1}\|\mathbf{s}_1 - \mathbf{s}_2\|_2\right)$. Proposition 6.1 yields consistency of the estimator under Type B or Type C sampling, as

$$E\|\widehat{C}_N - C\|_{\mathcal{S}}^2 \leq A\left(\exp(-2\rho^{-1}\rho_N) + I_{\rho_N}(\mathfrak{S}_N)\right).$$

If we assume a regular sampling design, then by Proposition 6.2

$$E\|\widehat{C}_N - C\|_{\mathcal{S}}^2 \leq A\left(\frac{1}{\alpha_N^d} + \frac{1}{N}\right).$$

Introducing the (unobservable) operator

$$\tilde{\Gamma}_N = \frac{1}{N}\sum_{k=1}^{N}(X(\mathbf{s}_k) - \mu) \otimes (X(\mathbf{s}_k) - \mu),$$

we see that

$$\tilde{\Gamma}_N - \widehat{\Gamma}_N = (\bar{X}_N - \mu) \otimes (\bar{X}_N - \mu).$$

Therefore

$$E\|\widehat{\Gamma}_N - C\|_{\mathcal{S}}^2 \leq 2E\|\tilde{\Gamma}_N - C\|_{\mathcal{S}}^2 + 2E\|(\bar{X}_N - \mu) \otimes (\bar{X}_N - \mu)\|_{\mathcal{S}}^2.$$

The bounds in Propositions 6.1, 6.2 and 6.3 apply to $E\|\tilde{\Gamma}_N - C\|_{\mathcal{S}}^2$. Observe that

$$E\|(\bar{X}_N - \mu) \otimes (\bar{X}_N - \mu)\|_{\mathcal{S}}^2 = E\|\bar{X}_N - \mu\|^4.$$

If $X(\mathbf{s})$ are bounded variables, i.e. $\sup_{t\in[0,1]}|X(\mathbf{s};t)| \leq B < \infty$ a.s., then $\|\bar{X}_N - \mu\|^4 \leq 4B^2\|\bar{X}_N - \mu\|^2$. It follows that under Assumption 3.1 we obtain the same order of magnitude for the bounds of $E\|\bar{X}_N - \mu\|^4$ as we have obtained in Propositions 5.1, 5.2 and 5.3 for $E\|\bar{X}_N - \mu\|^2$. In general $E\|\bar{X}_N - \mu\|^4$ can neither be bounded in terms of $E\|\bar{X}_N - \mu\|^2$ nor with $E\|\widehat{C}_N - C\|_{\mathcal{S}}^2$. To bound fourth order moments, conditions on the covariance between the variables $Z_{k,\ell} := \langle X(\mathbf{s}_{k,N}) - \mu, X(\mathbf{s}_{\ell,N}) - \mu \rangle$ and $Z_{i,j}$ for all $1 \leq i,j,k,\ell \leq N$ are unavoidable. However, a simpler general approach is to require higher order moments of $\|X(\mathbf{s})\|$. More precisely, we notice that for any $p > 1$, by the Hölder inequality,

$$E\|\bar{X}_N - \mu\|^4 \leq \left(E\|\bar{X}_N - \mu\|^2\right)^{1/p}\left(E\|\bar{X}_N - \mu\|^{\frac{4p-2}{p-1}}\right)^{(p-1)/p}.$$

Thus as long as $E\|X(\mathbf{s})\|^{\frac{4p-2}{p-1}} < \infty$, we conclude that, by stationarity,

$$E\|\bar{X}_N - \mu\|^4 \leq M(p)\left(E\|\bar{X}_N - \mu\|^2\right)^{1/p},$$

where $M(p)$ depends on the distribution of $X(\mathbf{s})$ and on p, but not on N. It is now evident how the results of Section 5 can be used to obtain bounds for $E\|\widehat{\Gamma}_N - C\|_{\mathcal{S}}^2$. We state in Proposition 6.4 the version for the general non-random design. The special cases follow, and the random designs are treated analogously. It follows that if Assumptions 3.1 and 3.2 hold, then $E\|\widehat{\Gamma}_N - C\|_{\mathcal{S}}^2 \to 0$, under Type B or C sampling, provided $E\|X(\mathbf{s})\|^{4+\delta} < \infty$.

PROPOSITION 6.4 *Let Assumptions 3.1 and 3.2 hold and assume that for some $\delta > 0$ we have $E\|X(\mathbf{s})\|^{4+\delta} < \infty$. Assume further that $\mathfrak{S}_N$ defines a non-random design of Type A, B or C. Then for any $\rho_N > 0$ we have*

$$(6.2) \quad E\|\hat{\Gamma}_N - C\|_{\mathcal{S}}^2 \leq 2\left\{H(\rho_N) + H(0)I_{\rho_N}(\mathfrak{S}_N)\right\} + 2C(\delta)\left\{h(\rho_N) + h(0)I_{\rho_N}(\mathfrak{S}_N)\right\}^{\frac{\delta}{2+\delta}}.$$

If $X(\mathbf{s}_1)$ is a.s. bounded by some finite constant B, then we can formally let δ in (6.2) go to ∞, with $C(\infty) = 4B^2$.

7 Inconsistent empirical functional principal components

We begin by formalizing the intuition behind Example 2.2. By Lemma 2.1, the claims in that example follow from Proposition 7.1. Recall that $X^\star = X(\mathbf{0}) \otimes X(\mathbf{0})$, and observe that for $x \in L^2$,

$$X^\star(x)(t) = \left(\int X(\mathbf{0}; u)x(u)du\right) X(\mathbf{0}; t) = \int c^\star(t, u)x(u)du,$$

where

$$c^\star(t, u) = X(\mathbf{0}; t)X(\mathbf{0}; u).$$

Since

$$E\iint \left(c^\star(t, u)\right)^2 dtdu = E\|X(\mathbf{0})\|^4 < \infty,$$

the operator $X^\star$ is Hilbert–Schmidt almost surely.

PROPOSITION 7.1 *Suppose representation (2.4) holds with stationary mean zero Gaussian processes ξ_j such that*

$$E[\xi_j(\mathbf{s})\xi_j(\mathbf{s} + \mathbf{h})] = \lambda_j\rho_j(h), \quad h = \|\mathbf{h}\|,$$
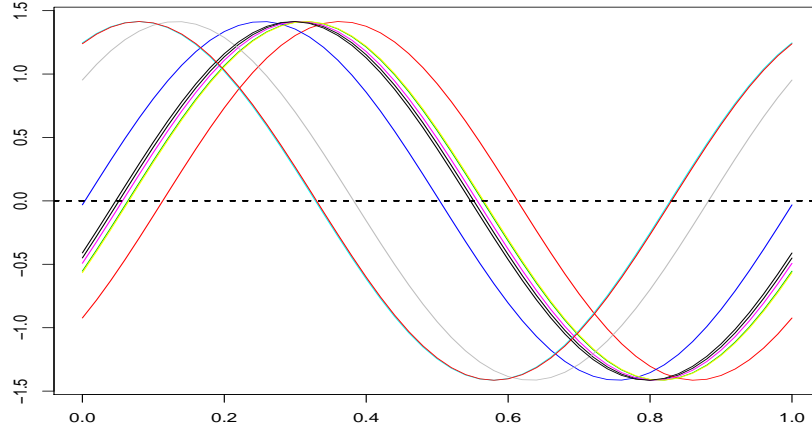
where each ρ_j is a continuous correlation function, and $\sum_j \lambda_j < \infty$. Assume the processes ξ_j and ξ_i are independent if $i \neq j$. If $\mathfrak{S}_N = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\} \subset \mathbb{R}^d$ with $\mathbf{s}_n \to \mathbf{0}$, then

$$(7.1) \qquad\qquad \lim_{N\to\infty} E\|\widehat{C}_N - X^\star\|_{\mathcal{S}}^2 = 0.$$

Proposition 7.1 is proven in Section 8.

We now present a very specific example that illustrates Proposition 7.1.

FIGURE 7.1 Ten simulated EFPC's $\hat{v}_1$ for process (7.2) with $\lambda = 0.5$ and $e_1(t) = \sqrt{2}\sin(2\pi t)$, $e_2(t) = \sqrt{2}\cos(2\pi t)$ ($N = 100$).

EXAMPLE 7.1 Suppose

$$(7.2) \qquad X(s;t) = \zeta_1(s)e_1(t) + \sqrt{\lambda}\zeta_2(s)e_2(t),$$

where the ζ_1 and ζ_2 are iid processes on the line, and $0 < \lambda < 1$. Assume that the processes ζ_1 and ζ_2 are Gaussian with mean zero and covariances $E[\zeta_j(s)\zeta_j(s+h)] = \exp\{-h^2\}$, $j = 1, 2$. Thus, each $Z_j := \zeta_j(0)$ is standard normal. Rearranging the terms, we obtain

$$X^\star(x) = \left(Z_1^2 \langle x, e_1 \rangle + \sqrt{\lambda}Z_1Z_2 \langle x, e_2 \rangle\right) e_1 + \left(\sqrt{\lambda}Z_1Z_2 \langle x, e_1 \rangle + \lambda Z_2^2 \langle x, e_2 \rangle\right) e_2.$$

The matrix

$$\begin{bmatrix} Z_1^2 & \sqrt{\lambda}Z_1Z_2 \\ \sqrt{\lambda}Z_1Z_2 & \lambda Z_2^2 \end{bmatrix}$$

has only one positive eigenvalue $Z_1^2 + \lambda Z_2^2 = \|X(0)\|^2$. A normalized eigenfunction associated with it is

$$(7.3) \qquad f := \frac{X(0)}{\|X(0)\|} = \left[Z_1^2 + \lambda Z_2^2\right]^{-1/2} \left(Z_1 e_1 + \sqrt{\lambda}Z_2 e_2\right).$$

Denote by $\hat{v}_1$ a normalized eigenfunction corresponding to the largest eigenvalue of $\widehat{C}_N$. By Lemma 2.1, $\hat{v}_1$ is close in probability to $\operatorname{sign}(\langle \hat{v}_1, f \rangle)f$. It is thus not close to $\operatorname{sign}(\langle \hat{v}_1, e_1 \rangle)e_1$.

Ten simulated $\hat{v}_1$, with $e_1(t) = \sqrt{2}\sin(2\pi t)$, $e_2(t) = \sqrt{2}\cos(2\pi t)$, $\lambda = 0.5$, are shown in Figure 7.1. The EFPC $\hat{v}_1$ is a linear combination of e_1 and e_2 with random weights. As formula (7.3) suggests, the function e_1 is likely to receive a larger weight. The weights, and so the simulated $\hat{v}_1$, cluster because both Z_1 and Z_2 are standard normal.

27

We now state a general result showing that Type A sampling generally leads to inconsistent estimators if the spatial dependence does not vanish.

PROPOSITION 7.2 *Assume that* $E\langle X(\mathbf{s}_1) - \mu , X(\mathbf{s}_2) - \mu\rangle \geq b(\|\mathbf{s}_1 - \mathbf{s}_2\|_2) > 0$, *where* $b(x)$ *is non-increasing. Then under Type A sampling the sample mean* $\bar{X}_N$ *is not a consistent estimator of* μ. *Similarly, if* $EX(\mathbf{s}) = 0$ *and*

$$(7.4) \qquad E\langle X(\mathbf{s}_1) \otimes X(\mathbf{s}_1) - C , X(\mathbf{s}_2) \otimes X(\mathbf{s}_2) - C\rangle_{\mathcal{S}} \geq B(\|\mathbf{s}_1 - \mathbf{s}_2\|_2) > 0,$$

where $B(x)$ *is non-increasing, then under Type A sampling the sample covariance* $\widehat{C}_N$ *is not a consistent estimator of* C.

We illustrate Proposition 7.2 with an example that complements Example 2.2 and Proposition 7.1 in a sense that in Proposition 7.1 the functional model was complex, but the spatial distribution of the $\mathbf{s}_k$ simple. In Example 7.2, we allow a general Type A distribution, but consider the simple model (7.2).

EXAMPLE 7.2 We focus on condition (7.4) for the FPC's. For the general model (2.4), the left–hand side of (7.4) is equal to

$$\kappa(\mathbf{s}_1, \mathbf{s}_2) = \sum_{i,j\geq 1} \mathrm{Cov}(\xi_i(\mathbf{s}_1)\xi_j(\mathbf{s}_1), \xi_i(\mathbf{s}_2)\xi_j(\mathbf{s}_2)).$$

If the processes ξ_j satisfy the assumptions of Proposition 7.1, then, by Lemma 2.2,

$$\mathrm{Cov}(\xi_i(\mathbf{s}_1)\xi_j(\mathbf{s}_1), \xi_i(\mathbf{s}_2)\xi_j(\mathbf{s}_2)) = \lambda_i^2 r_i + \lambda_j^2 r_j + \lambda_i\lambda_j\frac{r_i + r_j}{2} - \left(\lambda_i^{3/2} r_i + \lambda_j^{3/2} r_j\right) \sqrt{\lambda_i + \lambda_j},$$

where $r_i = \rho_i(\|\mathbf{s}_1 - \mathbf{s}_2\|)$.

To calculate $\kappa(\mathbf{s}_1, \mathbf{s}_2)$ in a simple case, corresponding to (7.2), suppose

$$(7.5) \qquad \lambda_1 = 1, \lambda_2 = \lambda, \ 0 < \lambda < 1, \quad \lambda_i = 0, \ i > 2, \quad \text{and} \quad \rho_1 = \rho_2 = \rho.$$

Then,

$$\kappa(\mathbf{s}_1, \mathbf{s}_2) = f(\lambda)\rho(\|\mathbf{s}_1 - \mathbf{s}_2\|),$$

where

$$f(\lambda) = (3 - 2\sqrt{2})(1 + \lambda^2) + 2\left[1 + \lambda + \lambda^2 - (1 + \lambda^{3/2})(1 + \lambda)^{1/2}\right].$$

The function f increases from about 0.17 at $\lambda = 0$ to about 0.69 at $\lambda = 1$.

We have verified that if the functional random field (2.4) satisfies the assumptions of Proposition 7.1 and (7.5), then $\widehat{C}$ is an inconsistent estimator of C under Type A sampling, whenever $\rho(h)$ is a nonincreasing function of h.

8 Proofs of the results of Sections 5, 6 and 7

PROOF OF PROPOSITION 5.1. By Assumption 3.1 we have

$$
E\left\|\frac{1}{N}\sum_{k=1}^{N}X(\mathbf{s}_{k,N})-\mu\right\|^{2}
$$

$$
=\frac{1}{N^{2}}\sum_{k=1}^{N}\sum_{\ell=1}^{N}E\langle X(\mathbf{s}_{k,N})-\mu\,,\,X(\mathbf{s}_{\ell,N})-\mu\rangle
$$

$$
\leq\frac{1}{N^{2}}\sum_{k=1}^{N}\sum_{\ell=1}^{N}h\big(\|\mathbf{s}_{k,N}-\mathbf{s}_{\ell,N}\|_{2}\big)
$$

$$
\leq\frac{1}{N^{2}}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\big(h(\rho_{N})I\{\|\mathbf{s}_{k,N}-\mathbf{s}_{\ell,N}\|_{2}\geq\rho_{N}\}+h(0)I\{\|\mathbf{s}_{k,N}-\mathbf{s}_{\ell,N}\|_{2}\leq\rho_{N}\}\big)
$$

$$
\leq h(\rho_{N})+h(0)\,I_{\rho_{N}}(\mathfrak{S}_{N}).
$$

∎

The following Lemma is a simple calculus problem and will be used in the proof of Proposition 5.2.

LEMMA 8.1 *Assume that f is a non-negative function which is monotone on $[0,b]$ and on $[b,\infty)$. Then*

$$
\sum_{k=0}^{L}f\left(\frac{k}{N}\right)\frac{1}{N}\leq\int_{0}^{L/N}f(x)dx+\frac{2}{N}\sup_{x\in[0,L/N]}|f(x)|.
$$

PROOF OF PROPOSITION 5.2. By Assumption 3.1,

$$
E\left\|\frac{1}{S_{N}}\sum_{k=1}^{S_{N}}X(\mathbf{s}_{k,N})-\mu\right\|^{2}=\frac{1}{S_{N}^{2}}\sum_{k=1}^{S_{N}}\sum_{\ell=1}^{S_{N}}E\langle X(\mathbf{s}_{k,N})-\mu\,,\,X(\mathbf{s}_{\ell,N})-\mu\rangle
$$

$$
\leq\frac{1}{S_{N}^{2}}\sum_{k=1}^{S_{N}}\sum_{\ell=1}^{S_{N}}h\big(\|\mathbf{s}_{k,N}-\mathbf{s}_{\ell,N}\|_{2}\big).
$$

Let $\mathbf{a}=(a_{1},\ldots,a_{d})$ and $\mathbf{b}=(b_{1},\ldots,b_{d})$ be two elements on $\mathcal{Z}(\boldsymbol{\delta})$. We define $d(\mathbf{a},\mathbf{b})=\min_{1\leq i\leq d}v_{i}(\mathbf{a},\mathbf{b})$, where $v_{i}(\mathbf{a},\mathbf{b})$ is the number of edges between a_{i} and b_{i}. For any two points $\mathbf{s}_{k,N}$ and $\mathbf{s}_{\ell,N}$ we have

(8.1) $\qquad\qquad d(\mathbf{s}_{k,N},\mathbf{s}_{\ell,N})=m \quad$ from some $m\in\{0,\ldots,KN^{1/d}\}$,

where K depends on $\mathrm{diam}(R_{0})$. It is easy to see that the number of points on the grid having distance m from a given point is less than $2d(2m+1)^{d}$, $m\geq 0$. Hence the number

29

of pairs for which (8.1) holds is $< 2d(2m+1)^{d-1}N$. On the other hand, if $d(\mathbf{s}_{k,N}, \mathbf{s}_{\ell,N}) = m$, then $\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2 \geq m\delta_0\eta_N$. Let us assume without loss of generality that $\delta_0 = 1$. Noting that there is no loss of generality if we assume that $x^{\delta-1}h(x)$ is also monotone on $[0, b]$, we obtain by Lemma 8.1 for large enough N and $K < K' < K''$

$$
\frac{1}{S_N^2} \sum_{k=1}^{S_N} \sum_{\ell=1}^{S_N} h\big(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,n}\|_2\big)
$$

$$
\leq 2d \sum_{m=1}^{K'N^{1/d}} \frac{(2m+1)^{d-1}}{N} h\big(m\eta_N\big) + \frac{2h(0)}{N}
$$

$$
\leq 2d\left(\frac{3}{\eta_N}\right)^{d-1} \sum_{m=0}^{K'N^{1/d}+1} \left(\frac{m}{N} N\eta_N\right)^{d-1} h\left(\frac{m}{N} N\eta_N\right)\frac{1}{N} + \frac{2h(0)}{N}
$$

$$
\leq 2d\left(\frac{3}{\eta_N}\right)^{d-1} \left(\int_0^{K''N^{1/d-1}} (N\eta_N x)^{d-1} h\big(N\eta_N x\big) dx \right.
$$

$$
\left. + \frac{2}{N} \sup_{x\in[0,K''\alpha_N/\Delta]} x^{d-1}h(x)\right) + \frac{2h(0)}{N}
$$

$$
= \frac{(3\Delta)^d d}{\alpha_N^d} \int_0^{K''\alpha_N/\Delta} x^{d-1}h\big(x\big) dx
$$

$$
+ \frac{4d(3\Delta)^{d-1}}{\alpha_N^{d-1}N^{1/d}} \sup_{x\in[0,K''\alpha_N/\Delta]} x^{d-1}h(x) + \frac{2h(0)}{N}.
$$

By Lemma 4.2, Type B sampling implies $\alpha_N \to \infty$ and $\alpha_N = o\big(N^{1/d}\big)$. This shows (5.3). Under Type C sampling $1/\alpha_N^d \ll 1/N$. The proof is finished. ∎

PROOF OF PROPOSITION (5.3). This time we have

$$
E\left\|\frac{1}{N} \sum_{k=1}^{N} X(\mathbf{s}_{k,N}) - \mu\right\|^2 = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\langle X(\mathbf{s}_{k,N}) - \mu,\, X(\mathbf{s}_{\ell,N}) - \mu\rangle
$$

$$
\leq \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} Eh\big(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2\big)
$$

$$
\leq \alpha_N^{-2d} \int_{R_N} \int_{R_N} h\big(\|\mathbf{s} - \mathbf{r}\|_2\big) f(\alpha_N^{-1}\mathbf{s}) f(\alpha_N^{-1}\mathbf{r})\, d\mathbf{s}d\mathbf{r} + \frac{h(0)}{N}
$$

$$
= \int_{R_0} \int_{R_0} h\big(\alpha_N\|\mathbf{s} - \mathbf{r}\|_2\big) f(\mathbf{s}) f(\mathbf{r})\, d\mathbf{s}d\mathbf{r} + \frac{h(0)}{N}.
$$

30

Furthermore, for any $\varepsilon_N > 0$,

$$\int_{R_0} \int_{R_0} h\big(\alpha_N \|\mathbf{s} - \mathbf{r}\|_2\big) f(\mathbf{s}) f(\mathbf{r}) \, d\mathbf{s} d\mathbf{r}$$

$$\leq h(0) \int_{R_0} \int_{R_0} f(\mathbf{s}) f(\mathbf{r}) I\big\{\|\mathbf{s} - \mathbf{r}\|_2 \leq \varepsilon_N\big\} \, d\mathbf{s} d\mathbf{r} + h(\alpha_N \varepsilon_N)$$

$$\leq \sup_{\mathbf{s} \in R_0} f^2(\mathbf{s}) \times \mathrm{Vol}\big\{(\mathbf{s}, \mathbf{r}) \in R_0^2 : \|\mathbf{s} - \mathbf{r}\|_2 \leq \varepsilon_N\big\} + h(\alpha_N \varepsilon_N).$$

∎

PROOF OF PROPOSITION 7.1. Observe that

$$\|\widehat{C}_N - X^\star\|_{\mathcal{S}}^2 = \iint \left\{ \frac{1}{N} \sum_{n=1}^N [X(\mathbf{s}_n; t) X(\mathbf{s}_n; u) - X(\mathbf{0}; t) X(\mathbf{0}; u)] \right\}^2 dt du.$$

Therefore,

$$\|\widehat{C}_N - X^\star\|_{\mathcal{S}}^2 \leq 2 I_1(N) + 2 I_2(N),$$

where

$$I_1(N) = \iint \left\{ \frac{1}{N} \sum_{n=1}^N X(\mathbf{s}_n; t)(X(\mathbf{s}_n; u) - X(\mathbf{0}; u)) \right\}^2 dt du$$

and

$$I_2(N) = \iint \left\{ \frac{1}{N} \sum_{n=1}^N X(\mathbf{0}; u)(X(\mathbf{s}_n; t) - X(\mathbf{0}; t)) \right\}^2 dt du.$$

We will show that $EI_1(N) \to 0$. The argument for $I_2(N)$ is the same. Observe that

$$I_1(N) = \frac{1}{N^2} \sum_{k,\ell=1}^N \iint X(\mathbf{s}_k; t)(X(\mathbf{s}_k; u) - X(\mathbf{0}; u)) X(\mathbf{s}_\ell; t)(X(\mathbf{s}_\ell; u) - X(\mathbf{0}; u)) dt du$$

$$= \frac{1}{N^2} \sum_{k,\ell=1}^N \int X(\mathbf{s}_k; t) X(\mathbf{s}_\ell; t) dt \int (X(\mathbf{s}_k; u) - X(\mathbf{0}; u))(X(\mathbf{s}_\ell; u) - X(\mathbf{0}; u)) du.$$

Thus,

$$EI_1(N) \leq \frac{1}{N^2} \sum_{k,\ell=1}^N \left\{ E\left(\int X(\mathbf{s}_k; t) X(\mathbf{s}_\ell; t) dt \right)^2 \right\}^{1/2} \left\{ E\left(\int Y_k(u) Y_\ell(u) du \right)^2 \right\}^{1/2},$$

where

$$Y_k(u) = X(\mathbf{s}_k; u) - X(\mathbf{0}; u).$$

We first deal with the integration over t:

$$E\left(\int X(\mathbf{s}_k; t) X(\mathbf{s}_\ell; t) dt \right)^2 \leq E \int X^2(\mathbf{s}_k; t) dt \int X^2(\mathbf{s}_\ell; t) dt$$

31

$$= E\left[\|X(\mathbf{s}_k)\|^2\|X(\mathbf{s}_\ell)\|^2\right] \le \left\{E\|X(\mathbf{s}_k)\|^4\right\}^{1/2}\left\{E\|X(\mathbf{s}_\ell)\|^4\right\}^{1/2} = E\|X(\mathbf{0})\|^4.$$

We thus see that

$$EI_1(N) \le \left\{E\|X(\mathbf{0})\|^4\right\}^{1/2}\frac{1}{N^2}\sum_{k,\ell=1}^{N}\left\{E\left(\int Y_k(u)Y_\ell(u)du\right)^2\right\}^{1/2}$$

$$\le \left\{E\|X(\mathbf{0})\|^4\right\}^{1/2}\frac{1}{N^2}\sum_{k,\ell=1}^{N}\left\{E\left(\int Y_k^2(u)du\right)^2\right\}^{1/4}\left\{E\left(\int Y_\ell^2(u)du\right)^2\right\}^{1/4}$$

$$= \left\{E\|X(\mathbf{0})\|^4\right\}^{1/2}\left[\frac{1}{N}\sum_{k=1}^{N}\left\{E\left(\int Y_k^2(u)du\right)^2\right\}^{1/4}\right]^2.$$

Consequently, to complete the verification of (7.1), it suffices to show that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\left\{E\left(\int Y_k^2(u)du\right)^2\right\}^{1/4} = 0.$$

The above relation will follow from

(8.2) $$\lim_{k\to\infty}E\left(\int Y_k^2(u)du\right)^2 = 0.$$

To verify (8.2), first notice that, by the orthonormality of the e_j,

$$\int Y_k^2(u)du = \sum_{j=1}^{\infty}\left(\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0})\right)^2.$$

Therefore, by the independence of the processes ξ_j,

$$E\left(\int Y_k^2(u)du\right)^2 = \sum_{j=1}^{\infty}E\left(\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0})\right)^4$$

$$+ \sum_{i\ne j}E\left(\xi_i(\mathbf{s}_k) - \xi_i(\mathbf{0})\right)^2 E\left(\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0})\right)^2.$$

The covariance structure was specified so that

$$E\left(\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0})\right)^2 = 2\lambda_j(1 - \rho_j(\|\mathbf{s}_k\|)),$$

so the normality yields

$$E\left(\int Y_k^2(u)du\right)^2 \le 12\sum_{j=1}^{\infty}\lambda_j^2(1 - \rho_j(\|\mathbf{s}_k\|))^2$$

$$+ 4\left\{\sum_{j=1}^{\infty}\lambda_j(1 - \rho_j(\|\mathbf{s}_k\|))\right\}^2.$$

32

The right hand side tends to zero by the Dominated Convergence Theorem. This establishes (8.2), and completes the proof of (7.1).

∎

PROOF OF PROPOSITION 7.2. We only check inconsistency of the sample mean. In view of the proof of Proposition 5.1 we have now the lower bound

$$E\left\|\frac{1}{N}\sum_{k=1}^{N}X(\mathbf{s}_{k,N})-\mu\right\|^2 \geq \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}b(\|\mathbf{s}_{k,N}-\mathbf{s}_{\ell,N}\|_2)$$
$$\geq b(\rho)I_\rho^2(\mathfrak{S}_N),$$

which is by assumption bounded away from zero for $N \to \infty$.

∎

References

Bel, L., Bar-Hen, A., Cheddadi, R. and Petit, R. (2010). Spatio-temporal functional regression on paleo–ecological data. *Journal of applied Statistics*, **00,** 000–000; Forthcoming.

Benko, M., Härdle, W. and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics*, **37,** 1–34.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer, New York.

Carroll, S. S. and Cressie, N. (1996). A comparison of geostatistical methodologies used to estimate snow water equivalent. *Water Resources Bulletin*, **32,** 267–278.

Carroll, S. S., Day, G. N., Cressie, N. and Carroll, T. R. (1995). Spatial modeling of snow water equivalent using airborne and ground–based snow data. *Environmetrics*, **6,** 127–139.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* Wiley.

Delaigle, A. and Hall, P. (2010). Defining probability density function for a distribution of random functions. *The Annals of Statistics*, **38,** 1171–1193.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21,** 224–239.

Du, J., Zhang, H. and Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Annals of Statistics*, **37,** 3330–3361.

Einmahl, J. H. J and Ruymgaart, F. G. (1987). The order of magnitude of the moments of the modulus of continuity of the multiparameter poisson and empirical processes. *Journal of Multivariate Analysis*, **21,** 263–273.

Gabrys, R., Horváth, L. and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, **105,** 1113–1125.

Gabrys, R. and Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, **102,** 1338–1348.

Gohberg, I., Golberg, S. and Kaashoek, M. A. (1990). *Classes of Linear Operators*. Operator Theory: Advances and Applications, volume 49. Birkhaüser.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components. *J. R. Statist. Soc* B, **68,** 109–126.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, **38,** 1845–1884.

Jiang, C-R. and Wang, J-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, **38,** 1194–1226.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparcity for principal components analysis in high dimensions. *Journal of the Americal Statistical Association*, **104,** 682–693.

Kaiser, M. S., Daniels, M. J., Furakawa, K. and Dixon, P. (2002). Analysis of particulate matter air pollution using Markov random field models of spatial dependence. *Environmetrics*, **13,** 615–628.

Lahiri, S. N. (1996). On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhya Series* A, **58,** 403–417.

Lahiri, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhya, Series* A, **65,** 356–388.

Lahiri, S. N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Annals of Statistics*, **34,** 1774–1813.

Lastovicka, J., Akmaev, R. A., Beig, G., Bremer, J., Emmert, J. T., Jacobi, C., Jarvis, J. M., Nedoluha, G., Portnyagin, Yu. I. and Ulich, T. (2008). Emerging pattern of global change in the upper atmosphere and ionosphere. *Annales Geophysicae*, **26,** 1255–1268.

Loh, W.-L. (2005). Fixed-domain asymptotics for a subclass of Matern-type Gaussian random fields. *Annals of Statistics*, **33,** 2344–2394.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2009). Removal of nonconstant daily variation by means of wavelet and functional data analysis. *Journal of Geophysical Research*, **114,** A03202.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2010a). Estimation of Sq variation by means of multiresolution and principal component analyses. *Journal of Atmospheric and Solar–Terrestial Physics*, **72,** 625–632.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2010b). Statistical significance testing for the association of magnetometer records at high–, mid– and low latitudes during substorm days. *Planetary and Space Science*, **58,** 437–445.

Park, B. U., Kim, T. Y., Park, J-S. and Hwang, S. Y. (2009). Practically applicable central limit theorem for spatial statistics. *Mathematical Geosciences*, **41,** 555–569.

Paul, D. and Peng, J. (2009). Consistency of restricted maximum likelihood estimators of principal components. *The Annals of Statistics*, **37,** 1229–1271.

Ramsay, J., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB.* Springer.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Springer.

Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, **102,** 984–996.

Rio, E. (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **29,** 587–597.

Rishbeth, H. (1990). A greenhouse effect in the ionosphere? *Planet. Space Sci.*, **38,** 945–948.

Roble, R. G. and Dickinson, R. E. (1989). How will changes in carbon dioxide and methane modify the mean structure of the mesosphere and thermosphere? *Geophys. Res. Lett.*, **16,** 1441–1444.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Krigging.* Springer.

Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case. *The Annals of Probability*, **12,** 361–379.

Ulich, T., Clilverd, M. A. and Rishbeth, H. (2003). Determining long-term change in the ionosphere. *Eos, Transactions American Geophysical Union*, **84,** 581–585.

Yamanishi, Y. and Tanaka, Y. (2003). Geographically weighted functional multiple regression analysis: A numerical investigation. *Journal of the Japanese Society of Computational Statistics*, **15,** 307–317.

Yao, F., Müller, H-G. and Wang, J-L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, **33,** 2873–2903.

Zhang, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99,** 250–261.